Руководство

ПО

конвертации бумажных словарей в Lingvo (для алчущих знаний пользователей)

Оглавление

<u>Введение</u>	3
<u>Глава 1. Поиск словаря.</u>	4
Глава 2. Сканирование, распознавание, сохранение.	5
Глава 3. Предфоматирование и вычитка	14
Быстрая вычитка.	27
Окончательная обработка текста и проверка сортируемых ошибок	
<u>Глава 4. Доводка и разметка словаря</u>	
<u>Приложения</u>	45
Замечания, соглашения и ЧЗВ (FAQ) – часто задаваемые вопросы	45
Название словарей и файлов словарей	45
Двухбуквенные сокращения названий языков	46
Список ссылок на сайты используемых программ	57
<u>ЧЗВ с форума Ru-Board, форума Lingvoda и мои собственные замечания.</u>	57
Создание и подключение иконки к словарю в Лингво.	69
<u>Работа с Hspell.</u>	72

Введение

Итак, после нескольких лет пользования Lingvo у многих возникает желание или жизненная потребность заиметь словарь, которого в Lingvo нет. Отбросив халявные пути, пойдем по пути более сложному – создадим словарь в Lingvo из бумажного словаря.

Глава 1. Поиск словаря.

Хорошо, когда требуемый бумажный словарь есть в наличии. Но что делать, когда его нет. Идем с паспортом в ближайшую крупную библиотеку – немного обаяния, пару шоколадок девочкам из читального зала, паспорт в залог и словарь Ваш (на время).

Если этот путь неприемлем или словаря в библиотеке нет – тогда надо бросать клич на Ru-Board, в форум Lingvo, форум Мультитрана, Города Переводчиков – мир не без добрых людей. Может кто-нибудь поможет.

Глава 2. Сканирование, распознавание, сохранение.

Дальше собственно его нужно отсканировать. На форуме Ru-Board, да и во многих местах в интернете есть много полезных советов и рекомендаций. У меня, после сканирования порядка 80 и конвертации порядка N-ного словарей выработались такие.

- Сканер лучше всего книжный (Пример Plustec OpticBook ~300 USD). У них один минус – сканируется не разворот, а страница, но из-за более высокой скорости время на сканирование всей книги не намного больше. К тому же такое сканирование гораздо менее губительно сказывается на книге, а для не своих словарей с клееным переплётом это единственный вариант отсканировать словарь, не разрушив его. Если не книжный, то обязательно ССD (ламповый). У них глубина резкости обычно достаточная, чтобы нормально отсканировать текст в зоне переплета. На мой взгляд, среди недорогих самые лучшие, быстрые и качественные это Epson (от ~ 80 USD).
- 2. Софт очень удобный скан-менеджер у FineReader. Он позволяет сканировать изображения сразу просматривая их на наличие дефектов, плохого сканирования разворота и т.п. Я пользуюсь им. Не рекомендую использовать 9 версию, поскольку открыть сохраняемые её изображения ничем другим я так и не смог, т.е. отсканированные ей изображения придется экспортировать.
- 3. Альтернатива сканирование в любой другой программе, допускающей пакетное сканирование (Scanitto, IrfanView и т.п.).
- 4. Как сканировать ч/б, серый, цветной. Данный вопрос довольно важен. Просто дам без комментариев свои рекомендации. Помните они НЕ догма.
 - Ч/б достаточно, только если книга хорошо разворачивается (т.е. развороты сканируются без черных полос, наползающих на буквы, все страницы книги отпечатаны качественно и бумага не просвечивается т.е. случай не часто встречающийся, причем старые словари в большинстве своем отпечатаны лучше новых). Яркость выставляйте такой, чтобы максимально убрать артефакты (обычно наиболее вредный артефакт пропечатка обратной стороны листа). Обычно хватает от 80 до 110 по Эпсоновской шкале. Если добиться приемлемого качества в данном интервале не удается, стоит переключиться в серый или цветной режим сканирования.
 - Серый обычно использую, если с ч/б не удается адекватно отсканировать разворот, печать бледная или проступают чернила с обратной стороны.
 - Цветной рекомендуется к использованию в редчайших случаях, если языки разделены цветом, очень плохая бумага и/или качество печати и времени не жалко.
- 5. Разрешение для простой латиницы и кириллицы при хорошем качестве печати достаточно 300 DPI, для немецкого, скандинавских и прочих языков, в алфавитах которых есть много навороченных символов с мелкими элементами, лучше 600 DPI. Что касается часто рекомендуемого разрешения 600 DPI то на простой латинице и кириллице при ответственном подходе к сканированию разница в уровне ошибок практически незаметна вычитывать все равно придется. Исключение словари с химическими, математическими и физическими формулами, словари с мелким текстом и на плохой бумаге.
- 6. Опять таки очень хороший вариант предложен в руководстве к Сканкромсатору Scan&Share 1.06 сканирование в сером на 300 DPI с конвертацией в ЧБ на 600 DPI и заодно разрезкой, обрезкой и выравниванием страниц.
- 7. За сканером при сканировании надо постоянно следить, удаляя со стекла попадающий из книги мусор (пыль, волосы и т.п.)
- 8. Как сканировать книгу разворотом или постранично.

Ответ: Если на сканер влазит разворот – сканируйте разворотом. Единственный момент – если жесткий корешок, что часто бывает у библиотечных словарей, которыми практически никто не пользовался – его надо хорошо разработать. Не бойтесь, старые словари имеют прочный сшитый потетрадный переплет, развалить который – это еще надо постараться. К сожалению, некоторые современные словари имеют проклеенный постраничный переплет, который часто разваливается даже при развороде в зоне переплета на 90 градусов. Их лучше не трогать, либо сразу договориться о переплете разваленного сканированием словаря.

Но самый лучший способ – если словарь свой и его не жалко – расшить переплет. В больших библиотеках есть ремонтные мастерские, где за пару рублей Вам переплетут словарь при необходимости обратно.

Итак, через пару часов мы отсканировали словарь.

Далее, при необходимости (читай отсутствии книжного сканера) крайне желательно обработать сканы, разрезать страницы, поудалять всю лишнюю информацию. Иногда полезно удалить колонтитулы и номера страниц. Это даст возможность впоследствии использовать автоматическое распознавание в Файнридере, а не наложение блоков в коррекцией там, где необходимо. В противном случае, удалять их из распознанного текста или еще в Файнридере частично придётся вручную. Для обработки отсканированных изображений страниц лучше всего подходит SkanKromsator.

Итак, мы получили отдельные красивые страницы. Просмотрите их. Если некотрые страницы имеют сильные искажения строк их необходимо исправить. Лучше всего с этим в автоматическом режиме справляется Finereader 9. Это, на мой взгляд, единственное действительное улучшение в данной версии.

Итак, импортируем полученные изображения в Файнридер, выставляем языки и жмем распознать всё.

6

Теперь нам необходимо обратить внимание на структуру словаря. В последнее время обычно словари имеют стандартную структуру, когда основное слово карточки идет с первой позиции строки, значения нумеруются по порядку, разные части речи, имеющие одинаковое значение и корневую основу, разделяются барабанными палочками «||», дополнительные примеры идут каждый с новой строки и основное слово в них заменяется тильдой. Текст на странице в зависимости от формата листа в 2 или 3 колонки.

```
carry-over 1. унос жидких или твёрдых
  частиц парами кипящей жидкости
  2. жидкие или твёрдые частицы, уне-
  сённые парами кипящей жидкости
cartridge:
 welding ~ флюсовая шайба
caryocerite кариоцерит (содержит то-
 рий)
cascade каскад (усиления)
 electrolytic ~ электролитический ка-
 скад, каскад ванн электролиза
case 1. оболочка; кожух; рубашка; обой-
 ма; ящик 2. (упрочнённый) поверх-
ностный слой (металла) 3. матрица
 (в электротипии)
  ~ of intense hardness поверхностный
 слой высокой твёрдости
 air \sim воздушная рубашка
 carburized ~ цементованный [наугле-
 роженный] поверхностный слой
 cyanided ~ цианированный поверх-
 ностный слой
```

Это самый лучший вариант для распознавания. Такие словари можно распознавать одновременно со сканированием. Только делать это все же лучше в фоновом режиме. У «Scan@Read» есть нехорошая фича, не позволяющая нормально просмотреть отсканированные страницы, пока открыто окно прогресса распознавания.

Но есть и другие варианты. В некоторых старых словарях словарные статьи пронумерованы.

```
apara teenaa ji on nonj tenas
43 angolo m di apertura anepryp-
    ный угол, ширина луча диа-
    граммы направленности.
44 angolo m di apertura del lobo
    di radiazione угол раствора
    лепестка излучения.
45 angolo m di arrivo угол при-
    хода (волны), угол падения.
46 angolo m di attacco угол ата-
    ки.
47 angolo m di campo угол зре-
    ния.
48 angolo m di circolazione угол
    отсечки.
49 angolo
               di
                   commutazione
           т
    угол перекрытия.
```

Если номера не были удалены в Skan Kromsatore теперь мы получим дополнительную кучу мусора, затрудняющую обработку. Такие словари потребуют дополнительно ручной

работы. Если они были распознаны на автомате – придется вручную перетаскивать границы блоков и перераспознавать блоки, чтобы номера остались за ними, либо, если поля на всех страницах одинаковые, вручную наносим блоки на одну страницу, сохраняем их, потом выделяем все остальные страницы, выбираем «Наложить блоки» и указываем сохраненный файл блоков.



Третий вариант словаря – когда вместо тильд в примерах используется сокращение из одной буквы.

```
аberwind теплый ветер с Альп, вы-
зывающий таяние снега
ability 1) способность, 2) характе-
ристика,
climbing a. характеристика ско-
роподъемности (самолета),
germinating a. всхожесть,
resolving a. разрешающая спо-
собность
ab initio лат сначала
```

Данный вариант – наихудший для последующей обработки, поскольку в данном случае интеллектуальный алгоритм распознавания FineReader (по сути, просто сверка со

встроенным словарем, либо алгоритм определения языка слова по сигнатурам, аналогичный использующемуся в Punto Switcher, Keyboard Ninja и т.п.) полностью беспомощен, и при распознавании, язык символов, имеющих одинаковое написание в кириллице и латинице, определяется, как Бог на душу положит (то же обычно получается и с распознаванием сокращений).

- 101 ability 1) способность, 2) характеристика, √
- 102 climbing a. характеристика скороподъемности (самолета), germinating a. всхожесть, resolving a. разрешающая способность↓
- 103 ab initio лат сначала↓
- 104 ablation абляция (уменьшение массы снега и

COBET

Можно после распознавания дополнительно сохранить результат в виде pdf-файла с текстом под изображением и при вычитке с использованием Adobe Acrobat пользоваться функцией поиска в последнем

И вот все грабли сканирования и распознавания позади.

Теперь, если словарь распознавался на автомате, желательно просмотреть его в FineReader еще раз и убедиться, что все распознано точно в соответствии с макетом страниц. Иногда FineReader проявляет собственную инициативу и при распознавании умудряется все же переделать блоки по своему, придав им конфигурацию напоминающую детские puzzles (никак не могу подобрать хороший русский перевод для данного слова в этом значении). Если такое есть - удаляем блоки, размечаем блок новой и распознаем его.

Итак, эпопея с распознаванием завершена. Возникает вопрос – стоит ли вычитывать в FineReader или выполнить эту работу в программе-редакторе? Вопрос имеет 2 ответа

- 1. ДА Если у Вас
 - от природы тяжело с грамотностью
 - алфавит языка словаря замороченный

- словарь специализированный, например химический с формулами или с ударениями на словах и эту информацию (формулы, ударения и т.п.) Вам хочется сохранить

- в словаре встречаются страницы с серьёзными дефектами печати

то, наверное, стоит вычитывать в FineReader. <u>Минусы</u> этого решения – Вы, скорее всего, окажетесь привязанными к одному компьютеру, да и скорость и удобство вычитки в FineReader будет не самым высоким. Да и потом придется предформатировать словарь для конвертации.

Плюсы – качество вычитки будет, наверное самым хорошим.

 HET - В программе-редакторе, параллельно с исправлениями предформатирования. <u>Плюсы</u> – получите файл практически готовый к конвертации, не будете привязаны к одному компьютеру. <u>Минусы</u> – качество вычитки, скорее всего, будет хуже, чем в варианте 1. Мы пойдем по варианту 2.

Итак, сохранение результатов – здесь сложностей немного, но есть один важный параметр.

Этот параметр – сохранение деления на строки.

Форматы	? ×
RTF/DOC/Word XML XLS PDF _HTM	IL PPT TXT DBF CSV LIT
Оформление: Таблицы, абзацы, шрифты	
Размер бумаги по умолчанию: А4	Автоматически увеличивать размер страницы
 Сохранять деление на строки Сохранять деление на страницы Сохранять цвет текста Удалять мягкий перенос 	 Выделять неуверенно распознанные символы Совместимость с Microsoft Word 95 Открывать окно Крупный план в Microsoft Word 2003 (только для формата Word XML)
Картинки Сохранять картинки Качество: Высокое (для печати)	Формат: (Авто)
	ОК Отмена Справка

Если мы сохраняем деление и удаляем мягкий перенос – некоторые места словаря превратятся в сплошной текст, но у нас практически не останется переносов в словах, при том, что сохранятся слова с дефисами.

Вот пример

displacement вытеснение; перемещение; сдвиг; 4 air d. перемещение воздуха, angle d. угловое смещение; angular d. см. angle displacement; 4 isentropic d. изэнтропическое перемещение [смещение]; phase d. фазовое смещение; сдвиг по фазе;↓ quasi-isentropic d. квазиизэнтро-пическое перемещение; seasonal trough d. сезонное перемещение ложбины; transisobaric d. of (the) air перемещение воздуха через изобары. transverse d. поперечное смещение; displacement of sunspots смещение солнечных

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

Если мы выберем сохранять деление на строки — все наоборот: в словаре все строки будут как в оригинале, благодаря чему мы сможем легко исправить тильды в начале строк, которые почему-то очень плохо распознаются FineReaderom, но частично разбивку на строки придётся исправлять вручную.

Форматы	? ×
RTF/DOC/Word XML XLS PDF HTML PPT TXT DBF CSV LIT	
Оформление: Таблицы, абзацы, шрифты	
Размер бумаги по умолчанию: А4 Автоматически увеличивать размер страницы	
Текст Сохранять деление на строки Сохранять деление на страницы Сохранять цвет текста Удалять мягкий перенос Картинки	ft 1L)
Сохранять картинки Качество: Формат: Высокое (для печати) (Авто)	-
ОК Отмена Справ	ка

Но в большинстве случаем, за исключением словарей со слишшком сложной структурой, с формулами и т.п. вариант 2 гораздо предпочтительнее.

Итак, сохраняем только информативные страницы словаря в Microsoft Word – как-то привычнее, или в текст Юникод.

Мастер сохранения результатов	Сохранить страницы	? ×
Выбрать приложение:	Папка: 😂 Мои документы 💽 🕓 🦻 🖡	୭
Сохранить страницы Image: Microsoft Word Microsoft Excel Image: Microsoft PowerPoint Image: Microsoft PowerPoint Image: WordPerfect StarWriter Image: Отправить страницы по электронной почте	[2004-02-09] в детском саду book 121 Corel User Files Adobe CyberLink Alcohol 120% DejaVu andrei DejavuX Automate_tasks detsad	Digi Dire Doc Doc Eps Exc
Форматы Какие страницы сохранять Выделенные страницы С Все страницы ОК Отмена	имя файла: П_КАЛОСХХ Тип файла: Текст Unicode (UTF-16) (*.txt) Сохранить: Эвыделенные страницы В Все страницы Сохранить как: Все страницы в один файл	Сохранить Отмена

Дополнительно сохраняем словарь в pdf с такими опциями.

А4	v
🔽 Сохранять размер оригинала	
Формат сохранения	
Текст под изображением страницы	
🗖 Сохранять цвет фона и букв	Paзрешить теги PDF (для Acrobat 5.0 или более поздней версии)
Картинки и шрифты	
Качество:	Формат:
Высокое (для печати)	(Авто)
Шрифт	
Использовать стандартные шрис	🛛 Встраивать шрифты
Защита	
Параметры защиты PDF Нет защи	иты

Пригодится для вычитки. Тем более и поиск в PDF будет работать.

Глава 3. Предфоматирование и вычитка

Открываем файл в редакторе. В каком редакторе производить данную работу – дело вкуса. Я предпочитаю EmEditor. Несмотря на громадный минус в виде отсутствия проверки орфографии (хотя на сайте есть плагины, позволяющие подцепить модуль проверки от Офиса, но в работе он крайне неудобен) плюсы этого редактора, особенно поддержка Регулярных выражений по Юникоду и возможность настройки с лихвой перекрывают этот большой минус.

И вот однородный черный текст – что может быть хуже для вычитки.



Поэтому нам надо создать свою конфигурацию, где бы все было бы не так печально.

Идем Menu-Configuration-Select Configuration-Define configuration-New. Выбираем Use default configuration и вводим имя новой конфигурации, например «En-Ru» (поскольку наш словарь англо-русский). Нам надо оптимально настроить программу для нашей работы.

На вкладке «General»

Wrap – Window width– будет переносить строки по ширине окна. Дополнительно отмечаем Show lines number (ну это понятно зачем).

Конвертация бумажных словарей в Lingvo

Свойства: New Configuration	<u>? ×</u>	
Highlight (1) Highlight (2) Display General Scroll File Backup A	Marks Print Link Keyboard Auto Save Association Wrap No Wrap 	
Wrap by:	Show Line Numbers	
Window Width	Show Page Numbers	
Line and Column Display as:	Show Ruler	
Logical Coord. (Full-Width Char. as 2 💌	🗖 Run Input Method Editor	
Normal Line Margin: 80 🚊	Allow Insert Control Character	
Quoted Line Margin: 🛛 80 😤	Save Insert/Overwrite Mode	
Quote Character: >	Preserve (CR/LF) returns on Clipboard	
[Undo] Max Number: 100000	Always Paste as ANSI No Space at Left Edge of Window	
	Tab/Indent Reset	
	ОК Отмена Справка	

Во вкладке «File» для удобства можем указать для сохранения кодировку UTF16-LE (Юникод).

На вкладке «Autosave» задайте интервал автосохранения.

На вкладке «Marks» отметим Show Returns, Show Tabs.

А теперь попытаемся настроить программу так, чтобы кириллица и латиница визуально отличались.

Переходим за вкладку «Highlight (1)». Жмем «Add» и вводим «[а-яё]» (чтобы не мучиться при вычитке словарей XX-РУ, подстраивая данный регекс под каждый язык). Ставим флажок «Regular expressions» и кнопкой F9 выбираем вид – оптимально, на мой взгляд, синий текст. (Дополнительная настройка на выделение для различия похожих символов, на типичные ошибки распознавания – это Вам домашнее задание). Недавно работая с немецко-русским словарем, в котором слова представлены в виде,

Вeschickungs || fahrzeug n самоходная загрузочная тележка (напр. с расходным бункером); ~förderer т питающий конвейср; ~gerät n загрузочное устройство; ~gut n загружасмый матернал; шихта; колоша, по-

чего лингвистический анализатор FineReader не понимает, (большое i) и (малое L) были часто перепутаны. Это конечно можно выделить и с помощью встроенных средств EmEditor, но в силу их ограниченности и нужности для более серьезных вещей, а также желания постоянно различать такую «петрушку» с любимым Arial Unicode MS немного пришлось подкорректировать данный шрифт. Буква Ы при этом тоже несколько изменилась (специалисты по шрифтам - исправьте эту граблю) но зато теперь «петрушка» хорошо видна.

logger: $ ightarrow$	\rightarrow \downarrow	
\rightarrow	radioactive ~ $ arrow$	прибор для рад <mark>иоа</mark> ктивного каротажа√
logging→	→ 1. каротах	₩ ↓
-	gamma <mark>-</mark> ray drill <mark>-</mark> hole	~ю́м. gamma <mark>-</mark> ray well logging√
\rightarrow	gamma <mark>-</mark> ray well ~→	гамма <mark>-</mark> каротаж скважины√
\rightarrow	neutron well ~ $ o$	нейтронный каротаж скважины√
\rightarrow	scintillation ~ $ o$	каротаж с помощью сцинтилляционного счётчика,

Далее раз уж в EmEditore проверки орфографии нет и ЗАО Информатик (<u>www.informatic.ru</u>) тоже не адаптирует свой замечательнейший продукт ОРФО (<u>www.orfo.ru</u>) под такие простые редакторы, попытаемся хотя бы сделать так, чтобы EmEditor показывал наиболее характерные ошибки распознавания текста. Ниже я приведу без объяснений листинг ESY-файла EmEditora с помощью которого Вам надеюсь удасться еще более улучшить результаты вычитки. Регексы здесь совсем простые, чтобы требовать каких-то комментариев.

Создано: 02.2008

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

```
EmEditor Syntax File
; To import this file to an existing configuration, select Configuration
 under the Tool menu, select Define Configurations, select a configuration
 you want to import to, press [Properties], select the Highlight page, and
 press [Import].
 Submit your customized file to submit@emurasoft.com
 to be listed in www.emurasoft.com user files pages.
#Highlight=on
#BeginTag=
#EndTag=
#CommentBegin=
#CommentEnd=
#LineComment1=
#LineComment2=
#SingleQuote=off
#DoubleQuote=off
#ContinueQuote=off
#Escape=
#ScriptBegin=
#ScriptEnd=
#SpecialSyntax=off
#HighlightBraces=off
#Keyword color=4,word=off,rightall=off,case=off,insidetag=off,regexp=on
[a-z]
#Keyword color=1,word=off,rightall=off,case=off,insidetag=off,regexp=on
#Keyword color=8,word=off,rightall=off,case=off,insidetag=off,regexp=on
(tlc|tlon|tn|rn|nlc|plc)
(и|п)(ая|ый|ий|ой|ое|ые|ия)
ы(п|н|ю)
([бвгдзжйклмнпрстфхцчшщьъ]{4,8})
([аеёиоуэюяьы]{3,6})
([aeiou]{4,8})
([bcdfghjklmnpqrstvwxz]{4,8})
ш(ыюшшыяжа)
([bcdfghjklmnpqrstvwxz])(l|f|j|t)([bcdfghjklmnpqrstvwxz])
(н|ь)(и|п)(ая|ый|ий|ой|ое|ые|ия)
```

Экспортируем его и жмем ОК и Close в окне списка конфигураций.

войства: En_Ru	?)
General Scroll File Backup Auto Save Highlight (1) Highlight (2) Display Marks	Association Wrap No Wrap Print Link Keyboard
✓ Highlight These Words	Add
Highlight Word	▲ Delete
 Іспінію) Іспінію Iспінію Iспінію	[F2] to Edit, [F9] to Toggle
8 🖊 ([аеёиоузюяьы]{3,6}) 8 🗸 ([aeiou]{4,8})	Whole Word Only
8 🖉 ([bcdfghjklmnpqrstvwxz]{4,8}) 8 🗸 ш(ыю)	Highlight Right Side Match Case
Image: Strain	🗾 🔲 Inside Tag Only
Begin Tag: End Tag:	Regular Expression
Import	Export Reset
ОК	Отмена Справка

Теперь, если Вы все сделали правильно, выбираем в Menu-Tools-Select Configuration вашу конфигурацию и получаем нечто вроде



Ну вот уже неплохо. В Menu-View настраиваем шрифт и размер (Font...).

Итак, первое, что мы сделаем - поубираем лишние пробелы. В силу того, что регулярные выражения (Regular Expressions) (далее RE) обрабатываются в EmEditor раз в 1000 медленнее, жмем Ctrl-H и в поле Find вводим два пробела, в поле Replace один пробел. Несколько раз нажмем ОК, и в конце внизу справа



Поскольку в дальнейшем наша основная работа в EmEditore будет производится именно с помощью Find/Replace, условимся о следующей форме записи:

F:XXXX//комментарий
R:YYYY
MC, RE
Что значит в поле Find ввести
«XXXX», в поле Replace – «YYYY»,
установить флажки Match Case (MC),
Use Regular Expressions (RE) (Use
Escape Sequence (ES)).

Replace	? X
Find:	<u>F</u> ind Next
	<u>R</u> eplace
Replace with:	Replace <u>A</u> ll
✓ Match <u>C</u> ase	Close
✓ Use Regular Expressions	
🔽 Use <u>E</u> scape Sequence	
🔲 Search Only <u>W</u> ord	
In the Selection Only	
Close when Finished	<u>H</u> elp

Итак, мы подошли к важнейшему моменту в создании словаря, от которого зависит количество труда, который мы потратим на дальнейшую обработку. Нам необходимо определиться со структурой текстового файла, которая бы при максимально легкой вычитке, также максимально легко позволила бы разделять словарные группы (имеется ввиду термин, перевод(ы), примеры, примечания). Этот довольно несложный вопрос для англо-русских словарей, но для словарей со сложной структурой, множеством различных помет и примечаний, как на языке перевода, так и на языке оригинала – правильность решения данного вопроса может сохранить Вам значительное количество времени при вычитке и предформатировании.

Моя схема для английских словарей такая

- 4 abating→ 1. снижение твёрдости (при отжиге наклёпанного материала)→ 2.
- удаление металла (обрубкой или обрезкой)√
- 5 aberration \rightarrow \rightarrow аберрация, отклонение \downarrow
- 6 → ~ arising from horizontal divergence → ошибка [погрешность], вызванная
- горизонтальной расходимостью (первичного пучка рентгеновских лучей)
- 7 → ~ arising from the use of flat plates → ошибка [погрешность], вызванная использованием плоской плёнки√

T.e.

Word [tab][tab] перевод 1[tab] перевод2 [tab] example[tab] перевод1[tab] перевод2

Для немецких Word [tab](род, число)[tab] [tab] перевод 1[tab] перевод2 [tab] [tab] example[tab] перевод 1[tab] перевод2

Для остальных – это Вам домашнее задание.

Соответственно, нам надо привести словарь к такой форме. Для примера я взял букву А из немецко-русского словаря по оптике.

При открытие его в Emeditor он будет выглядеть у Вас примерно так

кон<mark>-</mark>↓ 1 Abbe-Kondensor m 1. У нас имеются разбитые переносом денсор Аббе√ 2 русские слова. 3 Abbe-Kristall-Refraktometer 2. Грабли с распознаванием рода немецких п рефрактометр Аббе для↓ 4 слов. определения показателя \downarrow 3. Имеются различные мусорные символы ' 5 4. Тильда вначале часто распознана абы как. преломления в кристал-6 5. 7 лах√ 8 Abbe-Prisma n призма Аббе↓ Исходя из этого определяем стратегию Abbe<mark>-</mark>Refraktometer n peф<mark>-</mark>4 9 1. Удаляем множественные пробелы мусорные символы. рактометр Аббе 10 2. Исправляем тильды Abbe<mark>-Spektr</mark>ometer n спект<mark>-</mark>↓ 11 3. Объединяем разбитые русские слова и вообще переносим русские слова, рометр Аббе↓ стоящие 12 вначале строки. 13 abbeugen дифрагировать \downarrow 4. Исправляем род немецких слов Разбиваем текст табами. Abbeugung / 1. отклонение 5. 14 15 2. дифракция (см. также / 16 Beugung)↓ 17 Abbild n 1. изображение; \downarrow 18 отображение 2. снимок; 4 копия; отпечаток 3. кар<mark>-</mark>√ 19 тина (в плоскости навод<mark>-</mark>↓ 20 21 ки
системы) 22 4. модель (процесса)↓ 23 ~, gespiegeltes зеркальное↓ 24 отображение; зеркальное 25 изображение

И

Итак что же мы видим.

Поскольку у многих именно регексы вызывают трудности будем по возможности пользоваться ими, с параллельными комментариями.

Итак

1. Удаляем множественные пробелы

F:\+
R:
RE
Ищем пробел за ним еще как минимум один пробел — (\) пробел, + означает 1 и более.
201001501 US SPORTE DES MOY V KOOD OFFOO USU EQUAR (Core Oud USU Due) SODOUN

Заменяем на пробел. Для тех у кого Q6600 или E8xxx (Core Quad или Duo) первый пробел можно и убрать. Тогда будет заменять и одиночный пробел на одиночный пробел.

1а. Поскольку мусорные символы часто бывают разные к их замене надо подходить осторожно. Но здесь мы снова воспользуемся регексом

F:' ■
R:
RE
Ищем мусорные символы «'» или «∎». Символ означает ИЛИ.

После того, как ввели RE, нажмите FindNext, и если Emeditor найдет
нужный образец (он выделит его зеленой заливкой), ставьте галку In
the selection only (можете дополнительно выделить несколько строк
вокруг) и жмите ReplaceAll. Проверьте результат – то ли чего хотели?
Если Да - снимайте галку In the selection only и жмите ReplaceAll. Если
Нет - исправляйте исходный RE. (пример ниже)

2. Теперь, исправляем тильды. Здесь все гораздо сложнее, поскольку если мы посмотрим по всему файлу, то увидим, что тильда распознается как алфавитными так и пунктуационными символами. Поэтому частично сие придётся исправлять вручную.

F:^([^a-züöäßa-яё0-9]+), ([a-züöäß]+)

<u>R:∼, \2</u> RE

Ищем строки, начинающиеся (^ в самом начале регекса) с одного или более неалфавитных символов и цифр (все, что в квадратных скобках это требуемые символы – латиница а-z, умляуты и сцет, кириллица а-яё, цифры 0-9. ^ в квадратных скобках означает «исключая то, что в скобках») за которыми следует запятая, пробел и немецкое слово ([a-züöäß]+). В круглые скобки первое выражение можно не брать (тогда R: ~, \1), но для разделения групп искомых символов это лучше делать, тем более что ссылки (\1\2 и т.п. в R всегда относятся к содержимому круглых скобок)

Хоть данный регекс и дал результат, он не очень хороший, поскольку результат данного регекса сам является объектом поиска. И хотя здесь это не критично, т.к. результат меняется сам на себя, но это не всегда так. Поэтому максимально старайтесь избегать таких регексов.

2а. Теперь, исправляем тильды, оставшиеся после первого исправления, т.е. тильды при прямом порядке слов. Воспользуемся этим же самым макросом, только добавим запятую в список не искомых символов и уберём её между скобок и в Замене.

F:^([^a-züöäßa-яё0-9,]+) ([a-züöäß]+)	
R:~ \2	
RE	

3. Настало время объединить разбитые переносом русские слова.

F:([а-яё])\-\n([а-яё])
R:\1\2
RE, MC
Ищем русскую букву ([а-яё]) (в круглых скобках, чтобы использовать ссылки в Замене),
за ней дефис, новую строку (Для Emeditor \r\n и \n идентичны. Для некоторых других
редакторов - нет) и русскую букву в начале следующей строки. Используем «\n», а не
пытаемся выдумывать разную ерунду с «^», чтобы разорванную строку можно было
соединить. Заменяем \1(первая буква) \2(вторая буква) без всяких пробелов между.
Чувствительность к регистру (МС) – по желанию и потребностям.

Вот что мы должны получить

1	Abbe <mark>-</mark> Kondensor m конденсор Аббе√	- Kepter
2	Abbe <mark>-</mark> Kristall <mark>-</mark> Refraktometer↓	Find: [[a-яë]]\-
3	п рефрактометр Аббе для↓	Replace \1\2
4	определения показателя \downarrow	I Matc
5	преломления в кристаллах↓	🔽 Use f
6	<mark>Abbe</mark> -Prisma n призма Аббе√	Searce
7	Abbe <mark>-Refraktometer n рефрактометр Аббе√</mark>	In the
8	Abbe <mark>-</mark> Spe <mark>ktr</mark> ometer n спектрометр Аббе√	Close
9	abbeugen дифрагировать↓	
10	Abbeugung / 1. отклонение↓	
11	2. дифракция (см. также↓	
12	Beugung)↓	
13	Abbild n 1. изображение;↓	
14	отображение 2. снимок;↓	
15	копия; отпечаток 3. картина (в плоскости наводки оптической системы	ı)↓
16	4. модель (процесса)√	
17	~, gespiegeltes зеркальное \downarrow	

За. Теперь объединяем разбитые по строкам русские переводы. Самое простое – использовать практически тот же регекс, что и выше, только удалив дефис из F и добавив пробел в F. Однако при этом не будут объединены строки, разделённые по знаку препинания. Поэтому изменим данный регекс. Просмотрев пример мы увидим, что строки могут разделяться по «;»,«,»,«]»,«)» и т.п. В начале строки также может быть открывающая скобка.

F: <mark>([а-яё])</mark> ([\;\]\)\.]*)\n <mark>([\(\[]</mark> *)([а-яё])
R:\1\2 \3\4
RE
Ищем русскую букву ([а-яё]) за ней знак пунктуации, который может быть или нет (*

значит 0 и больше), перенос строки, возможно открывающую скобку и русскую букву. Надеюсь что замену пояснять не надо. 3b. Теперь осталось объединить переводы разделенные по значениям, причем так что цифра начинается с начала строки. Конечно это можно было добавить и в предыдущий регекс, но он стал бы слишком сложным. Поэтому слелаем это отдельно.

F:([а-яё])([\;\,\]\).]*)\n([0-9]{1,2})([\,\.]*) ([а-яё])

R:\1\2 \3. \5 RF

Ищем русскую букву ([а-яё]) за ней знак пунктуации, который может быть или нет (* значит 0 и больше), перенос строки, одну или две подряд цифры (значений переводов может быть и более 10), затем точку или запятую или ничего (запятая или ничего часто возникают из-за ошибок распознавания при плохом качестве оригинала), пробел и русскую букву. В Замене превращаем точку/запятую или их отсутствие просто в точку.

Зс . Теперь остались начинающиеся с начала строки переводы, полностью оторванные от исходного слова либо разбитые непонятными символами. Переносим их

F:\n([а-яё])	
R: \1	
RE	

Хе-хе. Это можно было использовать и вместо п. За. Но лишняя тренировка на начальном этапе освоения регексов никогда не помешает.

Итак получили это.

DE_O	ptic_aut1 ×		Beplace with:		
1	Abbe	Kondensor m конденсор Аббе√		Replace All	
-	A b b c		Match Case	Close	
2	eddA	Kristali-Refraktometer it pedpaktometp At	Use Regular Expressions	Help	
	препо	омпения в кристаплах↓	Use Escape Sequence		
			Search Only Word		
3	Abbe	Prisma n призма Аббе√	In the Selection Only		
Ŭ	/	norma n'inpriorita / to co t	Close when Finished		
4	Abbe <mark>-</mark>	Refraktometer n рефрактометр Аббе√ _ [↓]		111	
5	Abbe	Spektrometer n спектрометр Аббе√			

- 6 abbeugen дифрагировать
- 7 Abbeugung / 1. отклонение 2. дифракция (см. также√
- 8 Beugung)↓
- 9 Abbild n 1. изображение; отображение 2. снимок; копия; отпечаток 3. картина (в пл наводки оптической системы) 4. модель (процесса)↓
- 10 ~, gespiegeltes зеркальное отображение; зеркальное изображение↓
- 11 ~, getreues точное отображение; точное изображение
- 12 ~, räumliches объёмное изображение
- 13 abbilden 1. изображать; отображать 2. формировать изображение↓

По кириллице будем считать закончили. Теперь надо объединить разбитые переносом немецкие слова.

Делаем то же самое что и по кириллице, соответствующим образом исправив регекс

F:([a-züöäß])\-\n([a-züöäß])

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

R:\1\2

RE, MC

При этом помним, что разбитое (а не разделенное по дефису немецкое слово не может начинаться с большой буквы.

Итак это исправили.

Теперь осталось проверить строки. Начинающиеся непонятно с чего. Поскольку наш пример это только немецкие слова, начинающиеся с буквы А, это несколько упрощает задачу, хотя, поскольку данную проверку необходимо практически делать вручную, то на целых словарях, при переходе на другую букву просто вписывайте её в регекс и жмите Find Next.

|--|

RE,MC

Для некоторого облегчения данного процесса можно предварительно сдвинуть на табуляцию строки с примерами типа wort ~ - перевод.

Пожалуй мы так и сделаем.

:^([a-züöäß\]+) \~	
R:/t/O	
RE	

У нас таких оказалось всего одно.

Ну что ж, чуть модифицируем регекс выше и в путь

F:^([^AaäÄ\~\t])	
RE.MC	

Для данного случая лучше воспользоваться простой командой Найти (Find) чтобы по ошибке не удалить чего нибудь лишнее (если такое произошло нажмите Кнопку Отмена и последнее выполненное действие, даже регекс, будет отменено. Но это действительно не для всех редакторов). В зависимости от количества находимых однотипных ошибок может оказаться полезным остановиться, написать исправляющий часто встречающуюся ошибку регекс, прогнать его и продолжить исправление ошибок вручную. Попутно исправляем обнаруживаемые другие ошибки. Пару минут и файл проверен.

4. Исправляем род немецких слов. Пролистав словарь мы видим, что <u>f</u> чаще всего распознаётся как f,/,{, <u>m</u> – как m,т,т.,tri,tn,rn,ni,от,пг и т.п., а <u>n</u> – n,п,я,л. Также отметим, что они обычно разделяют слово и перевод, слово и номер значения перевода, сопровождаются указанием на число и могут через запятую отделять прилагательное устойчивого оборота.

Пишем регекс

F:([a-züöäß]) (f V \{ \}) ([\(\[]*) <mark>([1а-яё])</mark>
R:\1\t(f)\t\t <mark>\3\4</mark>
RE, MC
Ro amontus knyattus ekofikas uonoa aonmukattusko uonmy (oauauaiouusko utus) tonouuettam

Во вторых круглых скобках через вертикальную черту (означающую или) перечисляем все варианты. (f) – ну мне так нравится род в скобки заключать, а табы в замене сразу для предформатирования текста.

Повторяем для всевозмоных вариантов

F:([a-züöäß]) (f|V|\{|\}) pl([\.]*) (пробел в конце) R:\1\t(f) pl.\t\t

F:([a-züöäß]) (f|\/|\{|\})([\,\.]*) ([a-züöäß])

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

R:\1\t(f)\n\t\t~\3 \4 RE, MC

Повторяем все это и для т и п

F:([a-züöäß]) (m|т|т\.|tri|tn|rn|ni|от|пг) ([\(\[]*)([1а-яё]) R:\1\t(m)\t\t\3\4 RE,MC

F:([a-züöäß]) (m|т|т\.|tri|tn|rn|ni|от|пг) pl([\.]*) (пробел в конце) R:\1\t(m) pl.\t\t

F:([a-züöäß]) (m|т|т\.|tri|tn|rn|ni|от|пг)([\,\.]*) ([a-züöäß]) R:\1\t(m)\n\t\~\3 \4 RE, MC

F:([a-züöäß]) (n|п|я|л) ([\(\[]*)([1а-яё]) R:\1\t(n)\t\t\3\4 RE,MC

F:([a-züöäß]) (n|п|я|л) pl([\.]*) (пробел в конце) R:\1\t(n) pl.\t\t

F:([a-züöäß]) (n|п|я|л)([\,\.]*) ([a-züöäß]) R:\1\t(n)\n\t\t~\3 \4 RE, MC

Теперь осталось отделить наречия, глаголы, прилагательные и примеры от перевода

Наречия, глаголы, прилагательные

F:^([a-züöäß\(\)\-]+) ([\(\[]*)(<mark>[1а-яё]</mark>)	
R:\1\t\t\2\3	
RE	

Примеры

F:^\~(*) ([a-züöäß\(\)\-\ \[\]]+) ([\(\[]*)([1а-яё])	
R:\t\t~\1 \2\t\3\4	
RE	1

Примеры, удаленные табами от начала строки

F:^(\t+)([a-züöäß\(\)\-\ \[\]\~]+) ([\(\[]*)([1а-яё])	
R:\t/t\2\t\3\4	
RE	

Получили это

🕑 DE_U	ptic_auti " ×			Destruction and the second sec	
1	Abbe <mark>-</mark> Kondenso	or→	(f)→	→ KOHZehc Witt WW2/V3/4 > Replace All Replace All	
2	Abbe <mark>-</mark> Kristall <mark>-</mark> R	efrakton	neter→	(n) → → Match Case Uose Help NR	
	показателя пре	эломлен	ния в крі	ристаллах↓ Search Only Word	
3	Abbe <mark>-</mark> Prisma→	(n)→	\rightarrow	Призма Аббе Сове when Finished	
4	Abbe <mark>-</mark> Refraktor	neter→	(n)→	→ рефрактометр Аббе√	
5	Abbe <mark>-</mark> Spektrom	neter→	(n)→	→ спектрометр Аббе√	
6	abbeugen→	\rightarrow	\rightarrow	дифрагировать↓	
7	Abbeugung	(f)→	\rightarrow	1. отклонение 2. дифракция (см. также Beugung)↓	
8	Abbild→ (n)→	\rightarrow	1. изоб	бражение; отображение 2. снимок; копия; отпечаток 3.	
	картина (в пло	скости н	наводки	1 оптической системы) 4. модель (процесса)√	
9	\rightarrow \rightarrow	~, gesp	iegeltes [.]	узеркальное отображение; зеркальное изображение√	
10	\rightarrow \rightarrow	~, getre	eues⇒	точное отображение; точное изображение√	
11	\rightarrow \rightarrow	~, räum	nliches→	объёмное изображение√	
12	abbilden→	\rightarrow	\rightarrow	1. изображать; отображать 2. формировать изображение√	/
13	Abbildsbild	(n)→	\rightarrow	картина (в плоскости изображения оптической системы)↓	

Разумеется, в тексте будут ошибки предформатирования. Но их мы исправим позже в Excel.

Сейчас сохраняем отформатированный в Emeditor файл, а затем сохраняем еще раз в кодировке Utf-8 с BOM

<u>И</u> мя файла:	Untitled	•	Со <u>х</u> ранить
<u>Т</u> ип файла:	Text (*.bxt)	•	Отмена
Encoding:	Unicode		
	Add a Unicode Signature (BOM)		
Return Method:	No Change		

Быстрая вычитка.

Размышляя об оптимизации вычитки я давно пришел к выводу, что простая вычитка для словарей не оптимальный вариант. Надо как-то данный процесс автоматизировать. Но как??? Очень просто — взять все слова из файла и сравнить со словарями. Но где взять словари – да в Open Office. Вот так идея приобрела решение в мозгах, но как его реализовать, если в программировании не сильно бум. Благодарим <u>Andreyefg</u> который понял, оценил и самое главное реализовал эту идею практически, внеся в неё и свои полезные усовершенствония и идеи. Его мануал к программе, где работа программы объясняется чуть подробнее, приведен в приложении.

Запускаем программу Hspell. Щелкнем правой кнопкой мыши (ПКМ) в среднем окошке... и в контекстном меню выбираем пункт Get list of all words from text.



В открывшемся окне выбираем наш файл в кодировке Utf-8 с BOM



Программа автоматически создаст отсортированный с удалёнными дубликатами список слов, который будет показан в среднем поле.

H5pell 1.4.3				
Help				
Main				
				1
D Co a u	(фото)отпечаток			
	асимптотическое			
	22222			
	200			
T ru RU yo	Abathischer			
Backup-en_GB	Abbe-Kondensor			
Backup-en_US	Abbe-Kristall-Refraktometer			
en_GB	Abbe-Prisma			
en_US	Abbe-Refraktometer			
□ □ ru-ru_stems	Abbe-Spektrometer			
	abbeugen			
	Abbeugung			
	Applendung			
	Abbilden			
	applicen			
	Abbildung			
	Abbildungsapordnung			
	Abbildungsapertur			
	Abbildungsbedingung			
	Abbildungsbündel			
	Abbildungsdaten	-1		
, , , , , , , , , , , , , , , , , , , ,	2			
0		_	Select list	
		_	Select text Start	
	2962			

Общее количество слов в списке отображается в панели состояния (2962)

Теперь удалим из него слова имеющиеся в словарях. Словари у нас находятся в первом поле. Это словари из Open Office (пара файлов xx.aff и xx.dic) но в кодировке <u>UTF8!!!</u> (Стандартно ru_RU_yo словарь имеет кодировку KOI).

Выбираем необходимые словари, расставляя флажки.

HSpell 1.4.3		<u>- 🗆 ×</u>
Help		
Main		
		1
it_IT it_RU i	(фото)отпечаток /acummrotruveckoe /272721 200 Abathischer DAbbe-Kondensor Abbe-Kistall-Refraktometer Abbe-Ristall-Refraktometer Abbe-Refraktomete	
0 0		
	Select list	
	Select text Start	
	2962	

Затем щелкнаем правой кнопкой мыши (ПКМ) в среднем окошке и в контекстном меню выбираем пункт «**Remove correct spelled words**».

Получаем результат в строке состояния

HSpell 1.4.3	_ _ _ _ _
Help	
Main	
	1
t_TT ((boro)ornevarok Jackmiroruveckoe de_DE_neu Jackmiroruveckoe de_DE_neu Jackmiroruveckoe Jackbe-Rondensor Abbe-Kristal-Refraktometer Abbe-Kristal-Refraktometer Abbe-Prisma Abbe-Prisma Abbe-Spektometer Abbidungsbednun Abbidungsbednun Abbidungsbednun Abbidungsbedment Abbidungsbedment Abbidungsise Turru_stemet	ed 3
0 0 🔺 👘	
Select list	
Select text	Start
867	

Если установить флажок «Ignore case»

I/Pi	
/ец	✓ Ignore case
/ж	Save as
/Й	The Customize Font
1/22	

то перед проверкой все символы в словах будут переводиться в нижний регистр

Сохраним проверенный список в двухколоночный файл.

EHSpell 1.4.3	_ 🗆 🗵
Help	
Main	
Сарание и портисти (фото)отлечаток сарание и портисти и порти Портисти и портисти и поттисти и портисти и п	
□ de_DE_neu ????? □ ru_RU_je Abathischer □ ru_RU_yo □ Abbe-Kondensor	
Abbe-Kristall-Refraktometer Backup-en_US Abbe-Prisma en_GB Abbe-Refraktometer Abbe-Refraktometer Abbe-Refraktometer	
Abbergung Abbergung Abbergung Abbergung	
Abbildungsanordnun Abbildungsapertur Abbildungsbedingun Abbildungsbedingun Save as	
Abbildungsebene Abbildungsebene Abbildungselement Abbildungselement Abbildungserät Abbildungserät	
0 0	
Select list	
Select text Start	
867	



И откроем его в Emeditor.

Ľ	Example × [] DE_oplic_autos			
	überfokussierte	l→	überfok	ussierte↓
	übermikroskopis	sche→	\rightarrow	übermikroskopische↓
	unterfokussierte	→	\rightarrow	unterfokussierte↓
	unverkitteter⇒	Þ	unverki	tteter↓
	$\textbf{Wentzelsches} \\ \rightarrow$	\rightarrow	Wentze	lsches↓
	zentrierbare→	$\left \rightarrow \right.$	zentrier	bare↓
	zweilinsige∋	Þ	zweilins	sige↓
	Абата⇒∣⇒	Абата↓		
	Аббе→ ∣→	Аббе√		
	абберрациями	l→	абберр	ациями√
	авентуризирова	ать→	\rightarrow	авентуризировать√
	автокартограф	l→	автокар	отограф√
	автоколлимаци	онная→	\rightarrow	автоколлимационная√
	автоколлимаци	онный→	\rightarrow	автоколлимационный↓
	автомодуляция	H	\rightarrow	автомодуляция√

Исправлять слова необходимо в правой колонке, в исправляемые слова нельзя добавлять табуляции. (Такие ошибки исправляем вручную). Файл исходного списка и списка с частично исправленными ошибками приложены к данному документу. Тремя знаками вопроса, отмечены слова, которые потребовали ручного исправления.

После завершения исправлений сохраняем список в текст UTF8 с расширением hsp (его можно зарегистрировать в Emeditor как описано выше в настройке)

Возвращаемся в HSpell.



и выбираем сохранённый файл с исправлениями

C)ткрыть					? ×
	Папка:	1		•	🗢 🖻 💣 🎟 •	
	My Recent Documents Desktop My Documents My Computer	Example_corr.	hsp			
		Имя файла:	Example_corr		•	Открыть
	My Network Places	Тип файлов:	List of words (*.hsp)		•	Отмена
			ј Только чтение			

и выбираем отформатированный файл словаря (в

		spen 1.4.3
		Help
		Main
	e la	al
Открыть	<u>?</u> >	
		de_DE_neu : passed
Папка: 🗀 1	▼ 🗢 🖻 🚰 💷 -	ПП и RU /асимптотическое Пи_RU_yo: passed
		Current list was saved
DE Optic aut1		Abathischer
DE Optic aut 118		Abbe-Kondensor
		Backupen GB Abbe-Kristall-Refraktometer
My necerit DE_Optic_aut206		Backursen US Abbe-Prisma
Documents		Abbe-Betraktometer
		Abbe-Spektrometer
		Abbeugung Abbeugung
Deskton		Abbiendung
		Abbildshid
		Abbidungsapordnung
		Abhildumsanetur
		Abhildungsbadingung
My Documents		Abhidungsboangang
		Abhildungsdaten
		Abhildungsabana
		Abhildungselement
		Abhildungsteld
My Computer		A bhildungseta A bhildunggetät
· · · · · · · · · · · · · · · · · · ·		Abhildungsginsa
DE Optio aut1119	• Открыть	/ Vabilitariganitae
имя файла. [Dt_oplic_aurroo		Accepted: 854 Ignored: 13
Turn the waters: Tout file to process (* hit)	• Отмена	D:\mydoc\Example_corr.list Select list
My Network Thirt + Grades. Text file to process (.txt)		
Пасез Полько чтение		U-vnydociblackup-ub_upic_autr/U8.txt Select text Start
		867

Нажимаем кнопку

После этого сохраняем результат.



Внимание!!! Если вы введете имя уже существующего файла, он будет перезаписан Мы получили «вычитанный» файл

Помните!!!

- 1) Вычитываемый текст словаря должен иметь кодировку Utf-8 и только Utf-8.
- 2) Если что-то не работает, то первым делом проверяем везде кодировку.

Народ есть идея, добавив в программу поле редактора, сделав список двухколоночным для онлайн и оффлайн-проверки, добавив фукцию автоматического перехода на ошибку в тексте при перемещению по ошибкам в списке довести программу до ещё большего ума. Если кто желает помочь – обращаться к Andreyefg на форуме руборда или ко мне.

Окончательная обработка текста и проверка сортируемых ошибок

Теперь, поудаляв в нем лишние пробелы до и после табов, заменив двойные пробелы на одинарные и т.д. и т.п. импортируем файл в Эксель.

Мастер текстов (импорт) - шаг 2 и	из 3 Х
Данный диалог позволяет установить разбора.	» разделители для текстовых данных. Результат выводится в окне образца
Символом-разделителем является:	
I знак табуляции	
🗖 то <u>ч</u> ка с запятой	Считать последовательные разделители одним
<u>запятая</u>	
🗖 пробел	Ограничитель строк:
Другом.	
Образец разбора данн <u>ы</u> х	
Abbe-Kondensor	(f) конденсор Аббе
Abbe-Kristall-Refraktometer	r (n) рефрактометр Аббе для определения показателя 🔜
Abbe-Prisma	(n) призма Аббе
Abbe-Refraktometer	(n) рефрактометр Аббе
Abbe-Spektrometer	(n) спектрометр Аббе
	>
	Cancel < <u>Н</u> азад Дадее > <u>Г</u> отово
Мастер текстов (импорт) - шаг 3 г	из 3 🤶 🕺 🕺
Мастер текстов (импорт) - шаг 3)	
Мастер текстов (импорт) - шаг 3 г Данный диалог позволяет установить	из 3
Мастер текстов (импорт) - шаг 3 и Данный диалог позволяет установите ГФормат данных столбца	из 3 ? 🔀
Мастер текстов (импорт) - шаг 3 и Данный диалог позволяет установить Формат данных столбца С общий	из 3 ? X
Мастер текстов (импорт) - шаг 3 и Данный диалог позволяет установите -Формат данных столбца С общий "об откестовый осу	из 3 ? Х ь для каждого столбца формат данных. іщий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат -
Мастер текстов (импорт) - шаг 3 и Данный диалог позволяет установите Формат данных столбца Общий Отекстовый В д. В д.	из 3 ? Х Б для каждого столбца формат данных. щий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - ать, а всех прочих значений - в текст.
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установить Формат данных столбца Собщий Собщий Садта: ДМГ	из 3 ? Х ь для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подобнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установите Формат данных столбца С общий С дбщий С дата: ДМГ У С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. щий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установите Формат данных столбца Общий Сделий Сдата: ДМГ Спропустить столбец	из 3 ? Х ь для каждого столбца формат данных. щий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установитя Формат данных столбца С общий С дата: ДМГ С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. іщий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установити Формат данных столбца С общий С дата: ДМГ С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. іщий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установите Формат данных столбца Осщий С деший С дата: ДМГ С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. іщий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установити Формат данных столбца одбщий дата: ДМГ пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. іщий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установитя Формат данных столбца С общий С дата: ДМГ С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. іщий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установитя Формат данных столбца С общий С дата: ДМГ С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. ищий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установите Формат данных столбца Образи данных столбец С дата: ДМГ С пропустить столбец	из 3 ? Х 5 для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установитя Формат данных столбца Одинй Одинй С дата: ДМГ Опропустить столбец Образец разбора данных Собразец разбора данных Искот Rbbe-Knistall-Refraktometer	из 3 ? Х ь для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата шествляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установить Формат данных столбца С общий С дата: ДМГ С пропустить столбец Образец разбора данных С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее (f) конденсор Аббе х (n) рефрактометр Аббе для определения показателя (n) призиа Аббе
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установите Формат данных столбца Образи разбора данных С пропустить столбец Образец разбора данных Сбразец разбора данных	из 3 ? Х 5 для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее (f) конденсор Аббе x (n) призма Аббе (n) призма Аббе (n) призма Аббе
Мастер текстов (импорт) - шат 3 / Данный диалог позволяет установитя Формат данных столбца © общий © текстовый © дата: ДМГ © пропустить столбец Спропустить столбец Собразец разбора данных Собразец разбора данных Гекст Rbbe=Rondensor Rbbe=Rristall-Refraktometer Rbbe-Refraktometer	из 3 ? Х ь для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее и подробнее
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установитя Формат данных столбца С общий С дата: ДМГ С дата: ДМГ С пропустить столбец Образец разбора данных С пропустить столбец Сбразец разбора данных С пропустить столбец Сбразец разбора данных С дест столбец Сбразец разбора данных С пропустить столбец	из 3 ? Х ь для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее Какана и призика Аббе (n) рефрактометр Аббе (n) рефрактометр Аббе (n) рефрактометр Аббе
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установите Формат данных столбца © общий © текстовый © дата: ДМТ © пропустить столбец Образец разбора данных Спропустить столбец Собразец разбора данных Собразец разбора собра	из 3 ? Х 5 для каждого столбца формат данных. ший" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее (i) рефрактометр Аббе (n) рефрактометр Аббе (n) рефрактометр Аббе
Мастер текстов (импорт) - шат 3 и Данный диалог позволяет установите Формат данных столбца © общий © текстовый С дата: ДМГ © пропустить столбец Образец разбора данных Собразец разбора данных	из 3 ? Х 5 для каждого столбца формат данных. іщий" формат является наиболее универсальным. Для значений этого формата ществляется автоматическое преобразование числовых значений в числа, дат - аты, а всех прочих значений - в текст. Подробнее Тексетексетексет (1) конденсор Аббе (n) рефрактометр Аббе (n) рефрактометр Аббе (n) рефрактометр Аббе (n) рефрактометр Аббе (n) спектрометр Аббе

И жмем готово.

В итоге получаем в экселе 1 столбец – базовые слова, 2-й – род, 3 – примеры 4 – и т.д.

СРАЗУ ЖЕ выделяем первый столбец и жмем Меню-Вставка-Столбцы. У нас появляется пустой первый столбец. Прокрутим полосу прокрутки и посмотрим номер последней строки. Заполним 1 и 2 ячейки 1-го стобца 1 и 2 и далее Меню-Правка-Заполнить-Прогрессия – и в поле Предельное значение вводим номер последней строки словаря.

	A1	35	· (2) ·			_	-	-				
4	А	В	С	Расположение	Тип	E/	циницы •	1	J	к	L	M
1	1	Abbe-Ko	n (f)	С по строкам	• ар <u>и</u> фметическая	1	день					
2	2	Abbe-Kri	s (n)	In стольцам	С сеометрическая		рабочий день	теля	преломл	ения в кри	сталлах	
3		Abbe-Pri	s (n)		С		месяц					
4		Abbe-Re	fi(n)		втозаполнение		лод					
5		Abbe-Sp	e (n)	Автоматическое	определение шага							
6		abbeuge	n	Illar: 1	Предельное зна	чение:	1243	10				
7		Abbeugu	r (f)	Ear. 1.	npogoninico gila		3	ng)				
8		Abbild	(n)			ок	Отмена	ия; от	печаток	3. картина	(в плоскос	ти навод
9		1000		", gespieg зерк	альное отображен	ие; зе	ркальное изо	ражен	ие			
10				~, getreue точн	ое отображение; т	очное	изображение	2				
11				~, räumlic объё	мно <mark>е изображе</mark> ни	e						
12		abbilden		1. из	ображать; отображ	кать 2.	формировать	изобра	жение			
13		Abbildsb	il (n)	карт	ина (в плоскости из	вображ	кения оптичес	кой сис	темы)			
14		Abbildun	ε (f)	1. из	ображение; отобр	ажени	е (см. также В	ild) 2. п	роекция	изображе	ния)	
15				~, achrom axpo	матическое изобр	ажени	e					
16				~, afokale ador	альное изображен	ние						
17				~, anamor anaw	порфотное изображ	кение						
18				~, anastigraнас	тигматическое <mark>и</mark> зо	браже	ние					
19				~, aplanat апла	натическое изобра	жени	es					
20				~, asymptiacим	птотическое изобр	ражен	ие					
21				~, azimuta азим	утальная проекци	я						
22				~, brennpicм. А	bbildung, afokale							
23				~, detailre дета	льное изображени	ie; xop	ошо прорабо	танное	(фотограс	фическое)	изображе <mark>н</mark>	ние
24				~, dioptris диог	трическо <mark>е и</mark> зобра	жение						

Получаем нечто вроде

10	10	about-sled	lge	1. молот с 2. ковка сильными ударами						
11	11	abovegrou	nd	надземный						
12	12	abradant		см. abrasive						
13	13	abrade		1. истират 2. шлифовать; полировать; очищать						
14	14	abrader		машина для испытания истиранием						
15	15	abrading		притирка (напр. отверстия на валу)						
16	16	abrasion		1. истиран2. шлифовка; полировка; очистка						
17	17		oil ~	абразивное действие плохо очищенных масел						
18	18	abrasive		абразив, абразивный (шлифовальный) материал, истирающее вещество						
19	19		aluminum	естественный абразив, глинозёмный абразив, окись алюминия, применяє						
20	20		boride ~	борометаллический абразив						
21	21		boron nitra	боронитридный абразив						
22	22		coated ~	1. нанесён2. шкурка (шлифовальная или полировальная лента с нанесён						
23	23		metallic ~s	металлические абразивы						
24	24		silicon-car	карборундовый абразив						
25	25		synthetic ~	синтетические абразивы						

Нумерация крайне важна, поскольку нам она позволит восстанавливать исходный порядок слов после наших дальнейших операций. На всякий случай сохраним полученный файл в формате Эксель.

Начнем.

Выделяем весь лист щелчком по указанному квадрату.

	A	В	C	D	E	F	G	Н		
1	1	abac		номограм	ма					
2	2	abate		1. уменьц	2. отпуска	ть (сталь)				
3	3	aberration		отклонени	е, отступле	ение (напр.	от стандар	та)		
4	4		ability	способнос	ть					
5	5		~ to disma	способнос	ть к разбо	рке				
6	6		adhesive ~	способнос	ть к адгезі	и				
7	7		cutting ~	режущая	способнос	ть				
8	8		load carryi	несущая	способност	ъ, способн	ость выдер	живать		
9	9		running ~	эксплуата	ционные сі	войства				
10	10	about-sled	ge	1. молот с	2. ковка с	ильными у,	дарами			
11	11	abovegrou	nd	надземны	й					
12	12	abradant		см. abrasi	/e					
13	13	abrade		1. истират	2. шлифов	зать; полир	овать; очи	цать		
14	14	abrader		машина д	ля испытан	ния истиран	нем			
15	15	abrading		притирка (напр. отве	рстия на ва	алу)			
16	16	abrasion		1. истиран	2. шлифов	вка; полиро	вка; очисти	a		
17	17		oil ~	абразивно	е действие	е плохо очи	іщенных м	асел		
18	18	abrasive		абразив, абразивный (шлифовальный) материал, и						
19	19		aluminum	естественный абразив, глинозёмный абразив, оки						
20	20		boride ~	борометал	лический :	абразив				
21	21		boron nitra	боронитри	дный абра	зив				
22	22		coated ~	1. нанесён	2. шкурка	(шлифовал	льная или г	толирова		

Далее необходимо проверить правильность преформатирования. Меню-Данные-Сортировка-Столбец С потом В

	1	Calibri - 11 - A			С Общий	÷		=		B≁¤ B	ставить *	3			
E	Вставить	ж к ч - 🖾 - 👌 - 🗛			a - 🧐 - % 000	00 + 00	Условное	Форматир	овать Стили	3ª y	далить т				
Бyd	рер обмен	Б Шрифт	Б Выра	внивание	Га Число	G	рорматировани	е * как табли Стили	цу≁ ячеек≁	» ليون R	чейки	-			
	A1	• (* f* 1													
4	A	B	C (5)	Сортировка					1						
2		2 Abbe-Kristall-Refraktometer	(I) (n)	Арабави	гь уровень 🗙 Удалі	пъ уровени	- Сопиров	ать уровень	🍨 🏓 🖪	арамет	ры Г	Mo			
3		3 Abbe-Prisma	(n)	Столбец	III Creefer C	101	Сортировка				орядок				
4		4 Abbe-Refraktometer	(n)	Затен по	Стоябец В		Эначения				T A 80 9	_			
6		6 abbeugen	(11)		CTOREE D	<u>22</u>				<u>۱</u>	T HAV H				
7		7 Abbeugung	(f)												
8		8 Abbild	(n)												
10	1	0													
11	1	1		1											
12	1	2 abbilden 2 Abbildebild	(n)												
14	1	4 Abbildung	(f)					1. изобраз	жение; отоб	іражен	ие (см. та	кя			
15	1	5		~, ai	chromatische			ахромати	ческое изоб	ражен	ие				
16	1	6		~, at	okale			афокальн	ое изображе	ение	0	-			
18	1	8		, ai ~, ai	nastigmatische			анастигма	тическое из	обрая	ение				
19	1	9		~, aj	planatische			апланатич	ческое изобр	ражен	ие				
20	2	0		~, a:	ymptotische			асимптот	ическое изоб	браже	ние	_			
21	2	2		, a. ~, b	rennpunktlose			cm. Abbild	ыная проекці lung, afokale	ин					
23	2	3		~, d	etailreiche			детально	е изображен	ние; хо	орошо пр	op			
24	2	4		~, d	optrische			диоптрич	еское изобр	ажени	1e				
25	2	6		, d ~, e	ektronische			электрон	ощее изоора ное изображ	ажени кение	e				
27	2	7 Abbildung, emissionsmikrosko	pische изобра	жение, соз,	аваемое эмиссион	ным мик	роскопом, эм	иссионное	изображени	e					
2		164 aberrationsfrei											безаберра	ционны	Й
3		L66 Aberrationskoeffiz	ient m,												
1		171 abfangen											1. захваты	зать, ула	влива.
5		176 abfühlen											1. считыва	гь <mark>(н</mark> апр.	. перф
5	1	177 abgeblendet											задиафраг	мирован	ный (с
7	1	178 Abgleichlänge f													
3		182 abirren											рассеиват	ся1; отк	лонять
Э		197 Ablenkanordnung,	doppelts	vmmetr	ische отклоня	ющая	система	с двойно	ой симме	етри	ей				
D		215 Ablenkkondensato	prolatte f	, отклоня	ющая, пласти	ина. пл	пастина о	тклоняк	ошего ко	онле	нсатог	ba			
1		225 Ablenknuls in otkr	онающий	й импул	ьс										+
2		230 Ahlenksnannunger	molitudo	famo		alouter	о напреч								
5		59 Ablocoparallavo f	napaneou		сцёто	поще	о лаприя	sana							
2		209 Ableseparaliaxe,i	параллак	сприот	CHETE										
+		2/1 abreilsen											срывать (н	апр. ген	ераци
2		278 absaugen			_								отсасыват	, вытяги	івать (н
5		281 abschatten, abscha	ittieren I.	оттенят	ь 2. затенять										
7		285 abschirmen											экраниров	ать	
8	2	293 abschwächen											1. ослабля	гь; смягч	ать 2.
Э	3	300 absorbieren											абсорбирс	вать, по	глоща
D	3	314 Absorptionsgesetz	n, Bougu	ier-Lamb	ertsches зако	он Буге	ера—Лам	берта—	Бера						
1	:	323 Absorptionsquerso	hnitt in ce	ечение і	тоглощения										
2	:	324 Absorptionsspektr	alanalyse	, f абсор	бционный сп	ектра	льный ан	ализ							
3	:	331 abspiegeln											отражать		
1		345 abstrahlen											излучать		
5		352 abstreuen											рассеиват	,	
5		861 abtasten											nazeöntein	ath over	INDORC
		INT ADJACTELL											passepible	aro, undh	1/I/POBd

Вот сколько ошибок предформатирования. Исправляем их перемещая слова на свое место. При необходимости используем EmEditor (Копируем в EmEditor, прогоняем RE и вставляем обратно). Аналогично сортировкой по другим столбцам расставляем все на свои места попутно исправляя замеченные ошибки.

Трудноисправимые в таком виде ошибки выделяем цветом или специальным набором символов для последующего исправления в Emeditor.

Добавив еще один столбец в начало и использовав функцию Найти (запятую), потом выделяем весь столбец и специальной вставкой вставляем значения.

НАЙТИ	• (*	× ✓ fx =HA	<mark>іти(",";</mark> с	1;1)									A1		И(",";С1;1)
A	В	C D	E	F	G	Н	L	J	К	L	_		А	в с р	F F G
1 ,";C1;1)	1 Abb	e-Kon (f)		конденс	ор Аббе						1	#	ЗНАЧ!	Специальная вставка	<u>? ×</u>
2	2 Abb	e-Kris (n)		рефракт	ометр Аббе	для опре	еделения	показателя	преломл	ения в	2	#	знач!	Вставить	
3	3 Abl	Аргументы функ	ции							? ×	3	#	знач!	C BCE	○ с ис <u>х</u> одной темой
4	4 Abl	HOIATIA									4	#	знач!	C формулы	○ без рамки
5	5 Abl	HAPTPI				-	ਜ				5	#	знач!	• значения	С ширины столбцов
6	6 abt	Иско	чыи_текст	r 5"		E	9 = 5°				6	#	знач!	C форма <u>т</u> ы	C форму <u>л</u> ы и форматы чисел
7	7 Abl	Просматривае	ный_текст	r C1		Ē	= "Abbe	-Kondensor"			7	#	знач!	О приме <u>ч</u> ания	О значени <u>я</u> и форматы чисел
8	8 Abl	H	łaч_позиция	7 1		E	= 1				8	#	ЗНАЧ!	О условия на значения	
9	9						-			_	9	1 #	знач!	Операция	
10	10	Возвращает позици	но начала и	ккомой стро	ки текста в со	держащей (ее строке те	кста. Прописн	ые и строчн	ые		0 #	ЗНАЧ!	нет	С умножить
11	11	буквы различаютс	я.									1 #	3HAY!	С сло <u>ж</u> ить	С разделить
12	12 abt		Нач_г	юзиция по	зиция, с котор	юй нужно н	ачать поиск.	Первый знак	в параметр			2 #	значі	С вычесть	
13	13 Abl			УК	азано, началь	ная позиция	принимается	зя равной 1.	и эначение и			3 #	ЗНАЧТ		
14	14 Abl													🗖 пропускать пустые ячейки	тр <u>а</u> нспонировать
15	15	Значение:												1	
16	16	Справка по этой ф	7.38°11313					OK		a				Вставить связь	ОК Отмена
17	17	справка по этом ф	Y THE GRANT								LC	0)#	PARC		

Выделяем весь лист и сортируем по первому столбцу таким образом, чтобы заголовки (столбец С или третий) содержащие по два заголовка оказались вверху. (Можно воспользоваться фильтром, но если ошибок много фильтр не даёт копировать вставлять хуже

4	Α	В	С	D	E	F	G	н	1.00
1	#3HA4!	1	Abbe-Kon	(f)		конденсо	р Аббе		
2	#3HA4!	2	Abbe-Kris	(n)		рефракто	метр Абб	е для опре	еделения п
3	#3HA4!	Сортир	овка	_					
4	#3HA4!							1	
5	#3HA4!	°≩⊥До	об <u>а</u> вить уров	ень Хі	(далить уров	зень 🤖	сопировать у	ровень	🔄 🌵 🗖а
6	#3HA4!	Столб	eu			Сорти	оовка		
7	#3HA4!	Сортир	овать по	толбец А		• Значе	ния		
8	#3HA4!								-
9	#3HA4!								
10	#3HA4!								
11	#3HA4!								
12	#3HA4!								
13	#3HA4!								
14	#3HA4!								
15	#3HA4!								
16	#3HA4!								

8	593	Analyse, mikroschemische химический микроанализ			
9	982	Aufnahme, defokussierte расфокусированный снимок			
10	27	Abbildung, emissionsmikroskopische изображение, создава	емое	эмиссионным	микроскопом, эмиссионн
11	281	abschatten, abschattieren I. оттенять 2. затенять			
12	150	Abendkost f, Abendrot я,			
12	470	Adsorbens n, Adsorbent	(n)		адсорбирующее вещес
13	861	Atomtheorie ,f, Bohrsche теория атома Бора			
13	1152	Ausstrahlung, effektive эффективное излучение (земной п	оверх	ности)	
14	1119	Ausleuchten n, Ausleuchtung	(f)		высвечивание (люмин
15	1132	Ausschleusen n, Ausschleusung	(f)		шлюзование (объекта)
16	197	Ablenkanordnung, doppeltsymmetrische отклоняющая сист	ема с	двойной симме	етрией
16	436	Achsenversatz m, Achsenverschiebung f, Achsenversetzung	(f)		смещение (оптической
16	906	Auffängersystem, elektrisches система электрической реги	страц	ии (частиц)	
16	908	Auffangfläche /, Sharpsche светомерная пластинка Шарпа			
16	1036	Auftreffpunkt m, Auftreffstelle	(f)		точка (место) попадані
17	259	Ableseparallaxe ,f параллакс при отсчёте			
17	282	Abschattierung /, Abschattung	(f)		1. оттенок 2. оттенение
17	845	Atommittelpunkt ,т центр атома			
17	863	Atomverrückung f, Atomverschiebung	(f)		сдвиг [смещение] атом
17	1185	Auswahlprinzip n, Borsches принцип отбора Бора			
18	292	Abschlußfenster n, Abshluflglas n, Abschlußscheibe	(f)		защитное стекло
18	884	Auf dampfanlage /, Aufdampfeinrichtung	(f)		установка [устройство]
20	314	Absorptionsgesetz n, Bouguer-Lambertsches закон Бугера-	-Лам	берта—Бера	
22	421	Achsenhypermetropie f, Achsenhyperopie	(f)		осевая гиперметропия

Опять исправляем ошибки.

Аналогично с помощью сортировки по столбу с заголовками и сранению с порядком начальной нумерации можно найти ошибки в заголовках, влияющие на алфавитный порядок слов.

Ниже показан еще один пример ошибок из другого словаря
234		plain ~		см. cylindrical grir	nding		
579		dying ∼s		см. damped oscill	lations forced ~	вынужденные ко	лебания
191		cutoff ~		см. damper plate	1		
666		retarded ~		см. decelerated m	notion		
061		tooth ~		см. dedendum			
232		root ~		см. dedendum cir	cle slide valve ~	диаграмма Цейн	epa
934		root ~		см. dedendum line	e		
121	21579		dying ~s		см. damped os	cillations	
122	21579		forced ~s		вынужденные и	колебания	
123	23191		cutoff ~		см. damper plat	te	
124	20666		retarded ~		см. decelerated	motion	
125	8061		tooth ~		см. dedendum		
126	5232		root ~		см. dedendum (circle	
127	5232		slide valve ~		диаграмма Цей	інера	
100	47024					:	

Такие ошибки лучше исправлять вставкой специальной последовательности символов для последующей замены в Emeditor на новую строку или более сложный регекс.

<u>COBET</u>

Emeditor очень помогает в данной работе – выделяйте блоки, копируйте их в EmEditor и обрабатывайте с помощью RE. Потом обработанный материал вставляйте обратно.

Завершив все исправления, сортировкой по первому столбцу восстанавливаем исходный порядок, удаляем первый столбец, выделяем весь лист, копируем и вставляем в Emeditor или просто сохраняем файл в текст.

Регексом

F:\t+\$	
R:	
RE	

Удаляем табы в конце строки и заменяем все добавленные в экселе спецсимволы.

При необходимости снова проверяем файл в Экселе.

37

Глава 4. Доводка и разметка словаря

Итак, мы получили идеально предформатированный и вычитанный файл. Наши дальнейшие действия зависят того, как мы захотим увидеть примеры в нашем сконвертированном словаре. Есть два варианта.

- 1. Отдельная карточка + ссылка в теле основной карточки.
- 2. Пример в теле основной карточки.

Сделанный по первому способу словарь будет больше похож на оригинальный словарь от **Lingvo**, но и ошибок в нем из-за неправильного предфоматирования можно наделать поболее. Да и времени он потребует чуть поболе, размер словаря получится больше и по количественному составу гораздо ближе к тому, что заявляют создатели словарей, хотя иногда различия в количественном составе полученного Вами словаря и заявленное в бумажном оригинале могут Вас просто «убить». Но мы не ищем легких путей и пойдем по первому пути. Нам придется заменить тильды базовым словом.

Ниже текст макроса на VB, выполняющий данную процедуру. Предполагается, что примеров может быть не больше 1000. Желающие могут довести макрос до ума.

Sub Makpoc1()

```
'Макрос предназначен для замены тильд ~ на исходные слова - например в словарях.
Исходное слово д.б. в столбце 2 (b) словосочетание с тильдой в столбце 3 (c). 65500 –
лучше исправить на количество слов в вашем словаре
For i = 1 To 65500
If Cells(i, 2) <> "" Then baseword = Cells(i, 2)
For a = 1 To 1000
If Cells(i + a, 2) = "" Then
patchedword = Cells(i + a, 3): newpatchedword = Replace(patchedword, "~", baseword):
Cells(i + a, 3) = newpatchedword:
Else: GoTo 1
End If
Next a:
1: Next i
End Sub
```

Итак, мы пойдем по пути 1 и нам надо запустить макрос. Меню-Сервис-Макрос-Редактор Visual Basic двойной щелчок по Лист-Название нашего файла и копируем текст макроса в





Итак у нас получился финальный вычитанный файл в экселе.

	Α	В	С	D	E	F	G	Н	1	J	К	L	М	N 🗐
1		1 abac		номограмма										
2		2 abate		1. уменьшать, ослаблять (напр. силу, энергию)	2. отпуска	ть (сталь)								
3		3 aberration		отклонение, отступление (напр. от стандарта)										
4		4 ability		способность										
5		5	ability to dismantle	способность к разборке										
6		6	adhesive ability	способность к адгезии										
7		7	cutting ability	режущая способность										
8		8	load carrying ability	несущая способность, способность выдерживать	нагрузку;	грузоподъ	ёмность							
9		9	running ability	эксплуатационные свойства										
10		10 abac		номограмма										
11		11 abate		1. уменьшать, ослаблять (напр. силу, энергию)	2. отпуска	ть (сталь)								
12		12 aberration		отклонение, отступление (напр. от стандарта)										
13		13 ability		способность										
14		14	ability to dismantle	способность к разборке										
15		15	adhesive ability	способность к адгезии										
16		16	cutting ability	режущая способность										
17		17	load carrying ability	несущая способность, способность выдерживать	нагрузку;	грузоподъ	ёмность							
18		18	running ability	эксплуатационные свойства										
19		19 about-sledge		1. молот с поперечным задком; кувалда, кузнечн	2. ковка с	ильными у	дарами							
20		20 aboveground		надземный										
21		21 abradant		см. abrasive										
22		22 abrade		1. истираться, перетираться, срабатываться	2. шлифо	вать; поли	ровать; оч	ищать						
23		23 abrader		машина для испытания истиранием										
24		24 abrading		притирка (напр. отверстия на валу)										
25		25 abrasion		1. истирание, абразивный износ	2. шлифо	вка; полир	овка; очис	тка						
26		26	oil abrasion	абразивное действие плохо очищенных масел										
27		27 abrasive		абразив, абразивный [шлифовальный] материал	, истираю	цее вещес	тво абр	азивный,	шлифуюц	ий, истир	ающий			
28		28	aluminum oxide abras	естественный абразив, глинозёмный абразив, он	ись алюмі	ния, прим	еняемая д	іля полир	овки мета	лла				
29		29	boride abrasive	борометаллический абразив										
30		30	boron nitrate abrasive	боронитридный абразив										
31		31	coated abrasive	1. нанесённый (на подложку) абразивный пороц	2. шкурка	(шлифова	льная или	полирова	альная ле	нта с нане	сённым аб	разивом)		
32		32	metallic abrasives	металлические абразивы										
33		33	silicon-carbide abrasiv	карборундовый абразив										
34		34	synthetic abrasives	синтетические абразивы										
35		35	tumbling abrasives	абразивы, применяемые для полирования в гал	говочных (барабанах								
36		36 abrupt		1. резкий, отрывистый	2. внезап	ный								
37		37 abruption		обрыв, разрыв; поломка										
38		38 abscess		раковина (в металле)										
20 4 ♦ ▶	н	20 авсотв	ист3 / 🔁 /	1 послощать впитывать абсорбировать	2 300070	NOOD OT	1							

Первым делом с помощью сортировки по столбцу С из него получим такой.

	Α	В	С	D	E	F	G	Н	1	J	К	L	М	N	
13	13	ability		способность											
14	14		ability to dismantle												
15	15		adhesive ability												
16	16		cutting ability												
17	17		load carrying ability												
18	18		running ability												
19	19	about-sledge		1. молот с поперечным задком; кувалда, кузнеч	н 2. ковка с	ильными у	дарами								
20	20	aboveground		надземный											
21	21	abradant		см. abrasive											
22	22	abrade		1. истираться, перетираться, срабатываться	2. шлифов	вать; поли	ровать; оч	ищать							
23	23	abrader		машина для испытания истиранием											
24	24	abrading		притирка (напр. отверстия на валу)											
25	25	abrasion		1. истирание, абразивный износ	2. шлифов	вка; полир	овка; очис	тка							
26	26		oil abrasion												
27	27	abrasive		абразив, абразивный [шлифовальный] материа.	л, истираюь	цее вещес	тво абр	азивный,	шлифуюц	ций, истира	эющий				
28	28		aluminum oxide abrasi	ve											
29	29		boride abrasive												
30	30		boron nitrate abrasive												
31	31		coated abrasive												
32	32		metallic abrasives												=
33	33		silicon-carbide abrasiv	e											
34	34		synthetic abrasives												
35	35		tumbling abrasives												
36	36	abrupt		1. резкий, отрывистый	2. внезапи	ный									
37	37	abruption		обрыв, разрыв; поломка											
38	38	abscess		раковина (в металле)											
39	39	absorb		1. поглощать, впитывать, абсорбировать	2. аморти	зировать									
40	40	absorbent		поглотитель, абсорбент; гигроскопическое вещ	ество										
41	70	ability to dism	antle	способность к разборке											
42	71	ability to dism	antle	способность к разборке											
43	72	adhesive abili	ty	способность к адгезии											
44	73	adhesive abili	ty	способность к адгезии											
45	74	aluminum oxi	de abrasive	естественный абразив, глинозёмный абразив, о	кись алюми	ния, прим	иеняемая д	цля полир	овки мета	алла					
46	75	boride abrasiv	re	борометаллический абразив											
47	76	boron nitrate	abrasive	боронитридный абразив											
48	77	coated abrasiv	/e	1. нанесённый (на подложку) абразивный порог	и 2. шкурка	(шлифова	льная или	полиров	альная ле	нта с нанес	сённым аб	разивом)			
49	78	cutting ability		режущая способность											
50	79	cutting ability		режущая способность											
51	20 grozco abcorbont		ont	маслацый послотитоль											-

Теперь получим очень просто сформатированный словарь. Убеждаемся, что лист со словарем в книге называется Лист1. В противном случае в формулах ниже надо будет

подправить название листа.

На листе 2 в ячейке А1 вводим формулу

=ЕСЛИ(Лист1!В1<>"";СЦЕПИТЬ(Лист1!В1;"");СЦЕПИТЬ(Лист1!В1;""))

(для английской версии Экселя замените ЕСЛИ на IF и СЦЕПИТЬ на CONCATENATE) В ячейке В1 вводим формулу

=ЕСЛИ(Лист1!C1<>"";СЦЕПИТЬ("[ref]";Лист1!C1;"[/ref]");СЦЕПИТЬ(Лист1!C1;"")) В ячейках C,D,E и т.д. 1 вводим формулу

=ЕСЛИ(Лист1!D1<>"";СЦЕПИТЬ("[m1][trn]";Лист1!D1;"[/trn][/m]");СЦЕПИТЬ(Лист1!D1;""))

	A	В	C	D	E	F	-
13	ability		[m1][trn]способность[/trn][/m]				
14		[ref]ability to dismantle[/ref]					
15		[ref]adhesive ability[/ref]					
16		[ref]cutting ability[/ref]					
17		[ref]load carrying ability[/ref]					
18		[ref]running ability[/ref]					
19	about-sledge		[m1][trn]1. молот с поперечным задком; кувалда, кузнечный моло	[m1][trn]2. ковка сильными ударами[/trn][/m]			
20	aboveground		[m1][trn]надземный[/trn][/m]				
21	abradant		[m1][trn]см. abrasive[/trn][/m]				
22	abrade		[m1][trn]1. истираться, перетираться, срабатываться [/trn][/m]	[m1][trn]2. шлифовать; полировать; очищать[/t			
23	abrader		[m1][trn]машина для испытания истиранием[/trn][/m]				
24	abrading		[m1][trn]притирка (напр. отверстия на валу)[/trn][/m]				
25	abrasion		[m1][trn]1. истирание, абразивный износ [/trn][/m]	[m1][trn]2. шлифовка; полировка; очистка[/trn]			
26		[ref]oil abrasion[/ref]					
27	abrasive		[m1][trn]абразив, абразивный [шлифовальный] материал, истираю				
28		[ref]aluminum oxide abrasive[/ref]					
29		[ref]boride abrasive[/ref]					
30		[ref]boron nitrate abrasive[/ref]					
31		[ref]coated abrasive[/ref]					
32		[ref]metallic abrasives[/ref]					=
33		[ref]silicon-carbide abrasive[/ref]					
34		[ref]synthetic abrasives[/ref]					
35		[ref]tumbling abrasives[/ref]					
36	abrupt		[m1][trn]1. резкий, отрывистый [/trn][/m]	[m1][trn]2. внезапный[/trn][/m]			
37	abruption		[m1][trn]обрыв, разрыв; поломка[/trn][/m]				
38	abscess		[m1][trn]раковина (в металле)[/trn][/m]				
39	absorb		[m1][trn]1. поглощать, впитывать, абсорбировать [/trn][/m]	[m1][trn]2. амортизировать[/trn][/m]			
40	absorbent		[m1][trn]поглотитель, абсорбент; гигроскопическое вещество[/trn				
41	ability to dismantle		[m1][trn]способность к разборке[/trn][/m]				
42	ability to dismantle		[m1][trn]способность к разборке[/trn][/m]				
43	adhesive ability		[m1][trn]способность к адгезии[/trn][/m]				
44	adhesive ability		[m1][trn]способность к адгезии[/trn][/m]				
45	aluminum oxide ab		[m1][trn]естественный абразив, глинозёмный абразив, окись алюм				
46	boride abrasive		[m1][trn]борометаллический абразив[/trn][/m]				
47	boron nitrate abras		[m1][trn]боронитридный абразив[/trn][/m]				
48	coated abrasive		[m1][trn]1. нанесённый (на подложку) абразивный порошок [/trn][[m1][trn]2. шкурка (шлифовальная или полиро			
49	cutting ability		[m1][trn]режущая способность[/trn][/m]				
50	cutting ability		[m1][trn]режущая способность[/trn][/m]				
51	the Sucra Decorbont	2 (Berr3 / 9]	[m1][trn]woczausiń porzozurozs[/trn][/m]				

Размножаем строку на весь диапазон (все слова словаря) и радуемся. Выделяем весь лист, как описано

1	abac√								
2	\rightarrow	[m1][trn]номограмма[/trn][/r	m]→ →	\rightarrow	\rightarrow	\downarrow			
3	abate↓		-						
4	\rightarrow	[m1][trn]1. уменьшать, осла	аблять	(напр. с	силу, эн	нергию) [/t	rn][/m]→	[m1][tri	n]2.
	отпуска	ать (сталь)[/trn][/m]→ →	\rightarrow	\downarrow					
5	aberrat	on√			Repla	ice		<u>? ×</u>	
6	\rightarrow	[m1][trn]отклонение, отстуг	пление	(напр. с		t (m1)	Fin	d Next	
	станда	рта)[/trn][/m]→ → →	\rightarrow	\downarrow	Rep	lace with: fm11	Repl	Iace All	
7	ability↓					Match Case		llose	
8	\rightarrow	[m1][trn]способность[/trn][/	m] → →	\rightarrow		Use Regular Expressions Use Escape Sequence			
9	\rightarrow	[ref]ability to dismantle[/ref]-	\rightarrow \rightarrow	\rightarrow		Search Only Word In the Selection Only			
10	\rightarrow	[ref]adhesive ability[/ref]→	\rightarrow	\rightarrow		Close when Finished	<u> </u>	telp	
11	\rightarrow	[ref]cutting ability[/ref] $ o$ $ o$	\rightarrow	\rightarrow	\rightarrow	\downarrow			
12	\rightarrow	[ref]load carrying ability[/ref]	\rightarrow \rightarrow	\rightarrow	\rightarrow	\rightarrow	\downarrow		
13	\rightarrow	[ref]running ability[/ref]→→	\rightarrow	\rightarrow	\rightarrow	\downarrow			
14	abac√								
15	\rightarrow	[m1][trn]номограмма[/trn][/r	m]→ →	\rightarrow	\rightarrow	\downarrow			
16	abate↓								
17	\rightarrow	[m1][trn]1. уменьшать, осла	аблять	(напр. с	силу, эн	нергию) [/t	rn][/m]→	[m1][tri	n]2.
	отпуска	ать (сталь)[/trn][/m]→ →	\rightarrow	\downarrow					
вы	ше и к	опируем в EmEditor	, где р	регекс	сами -	гипа			
F: \	t\t[m1]								
R:\I	n\t\ [m 1]							
ES									
_									
F: \	[m1\]\	[trn\]([0-9]{1,2})\.) \\ пр	робел	і в кон	це				
R:\	[b\]\1.\	[/b\] \[m1\]\[trn\] \\ про	бел в	з конц	еим	ежду [/t	о] и [mʻ	1]	
RE									

COBET

Заменить (на ([i][com],) на [/com][i]), [на \[,] на \], см. хххх на [p]см.[/p] [ref]ххх[/ref] лучше сделать в соответствующих столбцах Экселя до вышеописанной процедуры (для предотвращения глупых вопросов как – через Emeditor). Применяя различные регексы к различным диапазонам можно быстро сформатировать довольно сложные по структуре словари.

доводим файл до ума.

Есть и еще вариант. После получения подготовленного к расстановке тэгов файла в Excel, расставим тэги на нестандартных примерах, а после, выделив весь словарь, перенесем его в Emeditor и обработаем нижеприведенным макросом. ВНИМАНИЕ !!! Данный макрос предназначен для конвертации только для англо-русских словарей. Для корректной работы с другими словарями требуется его адаптация. Скопируйте его в EmEditor и сохраните с названием MyMacro.jsee. Затем меню Macros-Select, Macros-Run

//Удаление пробелов перед табами и переносом строки document.selection.Replace("(\\+)\\t(\\+)","\\t",eeFindNext | eeFindReplaceEscSeq | eeReplaceAll | eeFindReplaceRegExp); document.selection.Replace("\\ +\$","".eeFindNext | eeFindReplaceEscSeg | eeReplaceAll | eeFindReplaceRegExp); document.selection.Replace("\\t+\$","",eeFindNext | eeFindReplaceCase | eeReplaceAll | eeFindReplaceRegExp);//Замена выражений типа (кириллица) на ([com][i]кириллица[/i][/com]) document.selection.Replace("(\\t|\\))\(([а-яё\\ \\-\\.\\,]+)\\)(\\ |\\;|\\,|\\.|\\t|\\n)","\\1[com]([i]\\2[/i])[/com]\\3",eeFindNext | eeFindReplaceEscSeq | eeReplaceAll | eeFindReplaceRegExp): document.selection.Replace("([а-яё])([\\,\\.\\ \\»]{1,2})\\(([^\\)\[]+)\\)","\\1\\2[com]([i]\\3[/i])[/com]",eeFindNext | eeFindReplaceEscSeq | eeReplaceAll | eeFindReplaceRegExp); //'Замена выражений типа [кириллица] на \[кириллица\] document.selection.Replace("(\\t|\\))\[([а-яё\\ \\-\\.\|,\\«\\»]+)\\](\\ |\\;|\\,|\\t|\\n)","\\1\\\[\\2\\\\]\\3",eeFindNext | eeReplaceAll | eeFindReplaceRegExp); //'Выделение pl. как пометы document.selection.Replace(" pl. "," [p]pl.[/p] ",eeFindNext | eeFindReplaceCase | eeReplaceAll); //Замена см. XXXX или см. также XXXX на [p]см. (тж.)[/p] [ref]XXX[/ref] document.selection.Replace("см\\. также ([a-z\\-\\]+)","[p]см. тж.[/p] [ref]\\1[/ref]",eeFindNext | eeReplaceAll | eeFindReplaceRegExp): document.selection.Replace("cm\\. ([a-z\\-\\]+)","[p]cm.[/p] [ref]\\1[/ref]",eeFindNext | eeReplaceAll | eeFindReplaceRegExp); // Замена \t XXXX на \t[ref]XXXX[[/ref] document.selection.Replace("^\\t([^\\[])([^\\t\\n]+)\\n","\\t[ref]\\1\\2[/ref]\\n",eeFindNext | eeFindReplaceEscSeg | eeReplaceAll | eeFindReplaceRegExp); // Замена 1-99. XXXX на [m1][b]1-99.[/b] [trn]XXXX[[/trn][/m] document.selection.Replace("\\t([0-9]{1,2})\\. ([^\\t]+)","\\n\\t[m1][b]\\1.[/b] [trn]\\2[/trn][/m]\\t",eeFindNext | eeFindReplaceEscSeq | eeReplaceAll | eeFindReplaceRegExp); // Замена \t\tперевод на \n\t[m1][trn]перевод[[/trn][/m] $document.selection.Replace("\t([^([])([^\t])),","\n(t[m1][trn])(1)(2[/trn][/m]),n",eeFindNext$ | eeFindReplaceEscSeq | eeReplaceAll | eeFindReplaceRegExp); // переводим на новую строку конструкции типа \t\t[p]см. (тж.)[/p] [ref]XXX[/ref] document.selection.Replace("\\t\\t([^\\n]+)\\n","\\n\\t[m1]\\1[/m]\\n",eeFindNext | eeFindReplaceEscSeg | eeReplaceAll | eeFindReplaceRegExp); // выделение несортируемой части английских глаголов document.selection.Replace("^to ","{to }",eeFindNext | eeReplaceAll | eeFindReplaceRegExp); // выделение кавычек как несортируемой части document.selection.Replace("^«([a-z\\-\\]+)» ","{«}\\1{»} ",eeFindNext | eeReplaceAll | eeFindReplaceRegExp); //Удаление табов в конце строки document.selection.Replace("\\t+\\n","\\n",eeFindNext | eeFindReplaceCase | eeReplaceAll | eeFindReplaceRegExp); //В начало документа и вставляем название словаря document.selection.StartOfDocument(false); document.selection.Text ="#NAME\t\"My first dictionary (En-Ru)\"\n"; document.selection.Text ="#INDEX_LANGUAGE\t\"English\"\n"; document.selection.Text ="#CONTENTS LANGUAGE\t\"Russian\"\n\n";

Сохраняем полученный dsl-файл и словарь почти готов. Требуются только мелкие доработки с учетом некоторых вопросов, освещённых в ЧЗВ (FAQ). Это домашнее задание.

Ну вот вроде и все, что хотелось добавить на данный момент. Я не скажу, что предложенные методы - лучшие, однако для простых пользователей они обладают серьезными преимуществами

- 1. Макросы в Emeditore позволяют легко и без особых навыков программирования автоматизировать данную работу с сохранением максимальной гибкости.
- 2. Использование Экселя позволяет без больших проблем создавать словари даже самой сложной структуры.

P.S. Будут вопросы – пишите. Постараюсь ответить – но работать Вашей головой я не буду. Если с помощью моих советов Вам удастся сконвертировать какой-нибудь словарь – присылайте, буду рад ;))))

Ниже еще кое-какая полезная информация, включая ЧЗВ.

Приложения

Замечания, соглашения и ЧЗВ (FAQ) – часто задаваемые вопросы

Чтобы упорядочить создание и распространение словарей под Лингво, на форуме Ruboard, где на данный момент и ведется основная работа по конвертации в Лингво словарных баз из существующего многообразия их источников, было принято несколько соглашений, касающихся названиях готовых словарей и некоторых других моментов.

Название словарей и файлов словарей

Имеется в виду название, которое указывается в начале файлов dsl - #NAME "Словарь". Далее это название отображается в списке словарей Лингво (Сервис - Словари или Ctrl+D), а также в верхней части словарной карточки.

Название словаря должно иметь вид: название (En-En)

где:

название - короткое название словаря (одно-два, максимум три слова). Примеры -Marketing, Словарь эпитетов, Big Soviet Encyclopedia. Тут главное, чтобы это название помещалось в списке словарей Лингво и по нему словарь легко было отличить от остальных словарей.

(En-En) - направление перевода словаря; для краткости должно быть представлено двумя двухбуквенными сокращенными названиями языков, между которыми стоит дефис. Ниже смотрите принятые сокращения самых распространенных языков.

Такое название словарей поможет с легкостью ориентироваться в них в Лингво.

Название файлов словарей

Файл словаря, скомпилированного под Лингво, должен иметь следующий вид:

En-En_abbr_автор_версия.zip (или .rar) - для файлов lsd En-En_abbr_автор_версия.dsl.zip (или .rar) - для файлов dsl

здесь:

En-En - пояснение см. выше

abbr - короткое название словаря (на латинице); максимум 3-4 слова. Кирилицу в названиях лучше не использовать - возможны проблемы с передачей файлов через интернет и другие трудности.

автор - сокращенное двухбуквенное название ника человека, скомпилировавшего этот словарь под Лингво. Ниже приведен список уже используемых сокращений. Указание авторства в названии файла не является обязательным.

версия - версия словаря вида v1.2. Предназначена для того, чтобы другим пользователям легче было разобраться - новый это словарь или старый, есть у них этот словарь или нет. Очевидно, когда словарь выкладывается впервые, ему присваивается версия 1.0 (если, конечно, это не пробная версия, предназначенная, скажем, только для ознакомления), при дальнейшей работе над словарем номер версии будет повышаться.

В названиях файлов не следует использовать апостроф (например, для обозначение мягкости согласных) - это может вызвать проблемы с передачей таких файлов, не все программы правильно обрабатывают этот знак. По той же причине в названиях не должно быть пробелов.

Файлы необходимо выкладывать либо в формате dsl (чтобы у пользователей старых версий Лингво было меньше трудностей с подключением данного словаря), либо и dsl, и lsd. В архиве вместе с файлом dsl должен также находится файл аннотации с расширением ann. Его содержание описано ниже.

Дополнительный текстовый файл

Вместе с архивом самого словаря должен присутствовать небольшой текстовый файл с таким же названием и расширением txt (т.е. En-En_abbr_автор_версия.txt). В этом файле должна содержаться следующая информация:

Название словаря (полное) кол-во заголовков - кол-во карточек Версия _____ от такого-то числа и года Источник Скомпилировано автор (данный пукт не обязателен) Куда обращаться при обнаружение ошибок - е-мейл, форум и т.п. (не обязательно)

Далее может содержаться информация об особенностях словаря, более полное его описание и другие данные, которые автор сочтет нужным сюда поместить. Можно, например, привести сокращенную (или полную) аннотацию из исходного словаря.

Первые обязательные строки (до источника включительно) должен также содержать файл аннотации (.ann). Автор также может включить в аннотацию любую другую информацию, которую сочтет уместной, в частности подобную той, что указана абзацем выше.

Двухбуквенные сокращения названий языков

Должны соответствовать стандарту ISO 639.1. Ниже полный перечень

ISO 639-	ISO	English name of Language
2 Code	639-1	
	Code	
aar	aa	Afar
abk	ab	Abkhazian
ace		Achinese
ach		Acoli
ada		Adangme
ady		Adyghe; Adygei
afa		Afro-Asiatic (Other)
afh		Afrihili
afr	af	Afrikaans
ain		Ainu
aka	ak	Akan
akk		Akkadian
alb/sqi	sq	Albanian
ale		Aleut
alg		Algonquian languages

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

alt		Southern Altai
amh	am	Amharic
ang		English, Old (ca.450-1100)
anp		Angika
apa		Apache languages
ara	ar	Arabic
arc		Aramaic
arg	an	Aragonese
arm/hye	hy	Armenian
arn		Mapudungun; Mapuche
arp		Arapaho
art		Artificial (Other)
arw		Arawak
asm	as	Assamese
ast		Asturian; Bable
ath		Athapascan languages
aus		Australian languages
ava	av	Avaric
ave	ae	Avestan
awa		Awadhi
aym	ay	Aymara
aze	az	Azerbaijani
bad		Banda languages
bai		Bamileke languages
bak	ba	Bashkir
bal		Baluchi
bam	bm	Bambara
ban		Balinese
baq/eus	eu	Basque
bas		Basa
bat		Baltic (Other)
bej		Beja
bel	be	Belarusian
bem		Bemba
ben	bn	Bengali
ber		Berber (Other)
bho		Bhojpuri
bih	bh	Bihari
bik		Bikol
bin		Bini; Edo
bis	bi	Bislama
bla		Siksika
bnt		Bantu (Other)
tib/bod	bo	Libetan
DOS	bs	Bosnian
bra		Braj
bre	br	Breton
btk		Batak languages
bua		Buriat
bug		Buginese

bul	bg	Bulgarian
bur/mya	my	Burmese
byn		Blin; Bilin
cad		Caddo
cai		Central American Indian (Other)
car		Galibi Carib
cat	са	Catalan; Valencian
cau		Caucasian (Other)
ceb		Cebuano
cel		Celtic (Other)
cze/ces	CS	Czech
cha	ch	Chamorro
chb		Chibcha
che	се	Chechen
chg		Chagatai
chi/zho	zh	Chinese
chk		Chuukese
chm		Mari
chn		Chinook jargon
cho		Choctaw
chp		Chipewyan
chr		Cherokee
chu	cu	Church Slavic; Old Slavonic; Church Slavonic; Old Bulgarian; Old
		Church Slavonic
chv	CV	Chuvash
chy		Cheyenne
cmc		Chamic languages
сор		Coptic
cor	kw	Cornish
COS	CO	Corsican
сре		Creoles and pidgins, English based (Other)
cpf		Creoles and pidgins, French-based (Other)
срр		Creoles and pidgins, Portuguese-based (Other)
cre	cr	Cree
crh		Crimean Tatar; Crimean Turkish
crp		Creoles and pidgins (Other)
csb		Kashubian
CUS		Cushitic (Other)
wel/cym	су	Welsh
cze/ces	CS	Czech
dak		Dakota
dan	da	Danish
dar		Dargwa
day		Land Dayak languages
del		Delaware
den		Slave (Athapascan)
ger/deu	de	German
dgr		Dogrib
din		Dinka
div	dv	Divehi; Dhivehi; Maldivian

doi		Dogri
dra		Dravidian (Other)
dsb		Lower Sorbian
dua		Duala
dum		Dutch, Middle (ca.1050-1350)
dut/nld	nl	Dutch; Flemish
dyu		Dyula
dzo	dz	Dzongkha
efi		Efik
egy		Egyptian (Ancient)
eka		Ekajuk
gre/ell	el	Greek, Modern (1453-)
elx		Elamite
eng	en	English
enm		English, Middle (1100-1500)
еро	eo	Esperanto
est	et	Estonian
baq/eus	eu	Basque
ewe	ee	Ewe
ewo		Ewondo
fan		Fang
fao	fo	Faroese
per/fas	fa	Persian
fat		Fanti
fij	fj	Fijian
fil		Filipino; Pilipino
fin	fi	Finnish
fiu		Finno-Ugrian (Other)
fon		Fon
fre/fra	fr	French
fre/fra	fr	French
frm		French, Middle (ca.1400-1600)
fro		French, Old (842-ca.1400)
frr		Northern Frisian
frs		Eastern Frisian
fry	fy	Western Frisian
ful	ff	Fulah
fur		Friulian
gaa		Ga
gay		Gayo
gba		Gbaya
gem		Germanic (Other)
geo/kat	ka	Georgian
ger/deu	de	German
gez		Geez
gil		Gilbertese
gla	gd	Gaelic; Scottish Gaelic
gle	ga	Irish
glg	gl	Galician
glv	gv	Manx

gmh		German, Middle High (ca.1050-1500)
goh		German, Old High (ca.750-1050)
gon		Gondi
gor		Gorontalo
got		Gothic
grb		Grebo
grc		Greek, Ancient (to 1453)
gre/ell	el	Greek, Modern (1453-)
grn	gn	Guarani
gsw		Swiss German; Alemannic
guj	gu	Gujarati
gwi		Gwich´in
hai		Haida
hat	ht	Haitian; Haitian Creole
hau	ha	Hausa
haw		Hawaiian
heb	he	Hebrew
her	hz	Herero
hil		Hiligaynon
him		Himachali
hin	hi	Hindi
hit		Hittite
hmn		Hmong
hmo	ho	Hiri Motu
scr/hrv	hr	Croatian
hsb		Upper Sorbian
hun	hu	Hungarian
hup		Нира
arm/hye	hy	Armenian
iba		Iban
ibo	ig	Igbo
ice/isl	is	Icelandic
ido	io	Ido
iii	ii	Sichuan Yi
ijo		ljo languages
iku	iu	Inuktitut
ile	ie	Interlingue
ilo		lloko
ina	ia	Interlingua (International Auxiliary Language Association)
inc		Indic (Other)
ind	id	Indonesian
ine		Indo-European (Other)
inh		Ingush
ipk	ik	Inupiaq
ira		Iranian (Other)
iro		Iroquoian languages
ice/isl	is	Icelandic
ita	it	Italian
jav	jv	Javanese
jbo		Lojban

jpn	ja	Japanese
jpr		Judeo-Persian
jrb		Judeo-Arabic
kaa		Kara-Kalpak
kab		Kabyle
kac		Kachin; Jingpho
kal	kl	Kalaallisut; Greenlandic
kam		Kamba
kan	kn	Kannada
kar		Karen languages
kas	ks	Kashmiri
geo/kat	ka	Georgian
kau	kr	Kanuri
kaw		Kawi
kaz	kk	Kazakh
kbd		Kabardian
kha		Khasi
khi		Khoisan (Other)
khm	km	Central Khmer
kho		Khotanese
kik	ki	Kikuyu; Gikuyu
kin	rw	Kinyarwanda
kir	ky	Kirghiz; Kyrgyz
kmb		Kimbundu
kok		Konkani
kom	kv	Komi
kon	kg	Kongo
kor	ko	Korean
kos		Kosraean
kpe		Kpelle
krc		Karachay-Balkar
krl		Karelian
kro		Kru languages
kru		Kurukh
kua	kj	Kuanyama; Kwanyama
kum		Kumyk
kur	ku	Kurdish
kut		Kutenai
lad		Ladino
lah		Lahnda
lam		Lamba
lao	lo	Lao
lat	la	Latin
lav	lv	Latvian
lez		Lezghian
lim	li	Limburgan; Limburger; Limburgish
lin	In	Lingala
lit	lt	Lithuanian
lol		Mongo
loz		Lozi

ltz	lb	Luxembourgish; Letzeburgesch
lua		Luba-Lulua
lub	lu	Luba-Katanga
lug	lg	Ganda
lui		Luiseno
lun		Lunda
luo		Luo (Kenya and Tanzania)
lus		Lushai
mac/mkd	mk	Macedonian
mad		Madurese
mag		Magahi
mah	mh	Marshallese
mai		Maithili
mak		Makasar
mal	ml	Malayalam
man		Mandingo
mao/mri	mi	Maori
map		Austronesian (Other)
mar	mr	Marathi
mas		Masai
may/msa	ms	Malay
mdf		Moksha
mdr		Mandar
men		Mende
mga		Irish, Middle (900-1200)
mic		Mi'kmaq; Micmac
min		Minangkabau
mis		Miscellaneous languages
mac/mkd	mk	Macedonian
mkh		Mon-Khmer (Other)
mlg	mg	Malagasy
mlt	mt	Maltese
mnc		Manchu
mni		Manipuri
mno		Manobo languages
moh		Mohawk
mol	mo	Moldavian
mon	mn	Mongolian
mos	-	Mossi
mao/mri	mi	Maori
may/msa	ms	Malay
mul		Multiple languages
mun		Munda languages
mus		Creek
mwl		Mirandese
mwr		Marwari
bur/mya	my	Burmese
myn		Mayan languages
myv		Erzya
nah		Nahuatl languages

nai		North American Indian
nap		Neapolitan
nau	na	Nauru
nav	nv	Navajo; Navaho
nbl	nr	Ndebele, South; South Ndebele
nde	nd	Ndebele, North; North Ndebele
ndo	ng	Ndonga
nds		Low German; Low Saxon; German, Low; Saxon, Low
nep	ne	Nepali
new		Nepal Bhasa; Newari
nia		Nias
nic		Niger-Kordofanian (Other)
niu		Niuean
dut/nld	nl	Dutch; Flemish
nno	nn	Norwegian Nynorsk; Nynorsk, Norwegian
nob	nb	Bokmål, Norwegian; Norwegian Bokmål
nog		Nogai
non		Norse, Old
nor	no	Norwegian
nqo		N'Ko
nso		Pedi; Sepedi; Northern Sotho
nub		Nubian languages
nwc		Classical Newari; Old Newari; Classical Nepal Bhasa
nya	ny	Chichewa; Chewa; Nyanja
nym		Nyamwezi
nyn		Nyankole
nyo		Nyoro
nzi		Nzima
oci	ос	Occitan (post 1500); Provençal
oji	ој	Ojibwa
ori	or	Oriya
orm	om	Oromo
osa		Osage
OSS	OS	Ossetian; Ossetic
ota		Turkish, Ottoman (1500-1928)
oto		Otomian languages
paa		Papuan (Other)
pag		Pangasinan
pal		Pahlavi
pam		Pampanga
pan	ра	Panjabi; Punjabi
рар		Papiamento
pau		Palauan
рео		Persian, Old (ca.600-400 B.C.)
per/fas	fa	Persian
phi		Philippine (Other)
phn		Phoenician
pli	pi	Pali
pol	pl	Polish
pon		Pohnpeian

por	pt	Portuguese
pra		Prakrit languages
pro		Provençal, Old (to 1500)
pus	ps	Pushto
qaa-qtz		Reserved for local use
que	qu	Quechua
raj		Rajasthani
rap		Rapanui
rar		Rarotongan; Cook Islands Maori
roa		Romance (Other)
roh	rm	Romansh
rom		Romany
rum/ron	ro	Romanian
rum/ron	ro	Romanian
run	rn	Rundi
rup		Aromanian; Arumanian; Macedo-Romanian
rus	ru	Russian
sad		Sandawe
sag	sg	Sango
sah		Yakut
sai		South American Indian (Other)
sal		Salishan languages
sam		Samaritan Aramaic
san	sa	Sanskrit
sas		Sasak
sat		Santali
scc/srp	sr	Serbian
scn		Sicilian
SCO		Scots
scr/hrv	hr	Croatian
sel		Selkup
sem		Semitic (Other)
sga		Irish, Old (to 900)
sgn		Sign Languages
shn		Shan
sid		Sidamo
sin	Si	Sinhala; Sinhalese
SÍO		Siouan languages
sit		Sino-Tibetan (Other)
sla		Slavic (Other)
slo/slk	sk	Slovak
slo/slk	SK	Slovak
SIV	SI	Siovenian
sma		Southern Sami
sme	se	Northern Sami
smi		Sami languages (Other)
smj		
smn		Inari Sami
smo	sm	Samoan
sms		Skolt Sami

sna	sn	Shona
snd	sd	Sindhi
snk		Soninke
sog		Sogdian
som	SO	Somali
son		Songhai languages
sot	st	Sotho, Southern
spa	es	Spanish; Castilian
alb/sqi	sq	Albanian
srd	SC	Sardinian
srn		Sranan Tongo
scc/srp	sr	Serbian
srr		Serer
ssa		Nilo-Saharan (Other)
SSW	SS	Swati
suk		Sukuma
sun	su	Sundanese
sus		Susu
sux		Sumerian
swa	SW	Swahili
swe	SV	Swedish
syr		Syriac
tah	ty	Tahitian
tai		Tai (Other)
tam	ta	Tamil
tat	tt	Tatar
tel	te	Telugu
tem		Timne
ter		Tereno
tet		Tetum
tgk	tg	Tajik
tgl	tl	Tagalog
tha	th	Thai
tib/bod	bo	Tibetan
tig		Tigre
tir	ti	Tigrinya
ti∨		Tiv
tkl		Tokelau
tlh		Klingon; tlhIngan-Hol
tli		Tlingit
tmh		Tamashek
tog		Tonga (Nyasa)
ton	to	Tonga (Tonga Islands)
tpi		Tok Pisin
tsi		Tsimshian
tsn	tn	Tswana
tso	ts	Tsonga
tuk	tk	Turkmen
tum		Tumbuka
tup		Tupi languages

tur	tr	Turkish
tut		Altaic (Other)
t∨l		Tuvalu
twi	tw	Twi
tyv		Tuvinian
udm		Udmurt
uga		Ugaritic
uig	ug	Uighur; Uyghur
ukr	uk	Ukrainian
umb		Umbundu
und		Undetermined
urd	ur	Urdu
uzb	uz	Uzbek
vai		Vai
ven	ve	Venda
vie	vi	Vietnamese
vol	VO	Volapük
vot		Votic
wak		Wakashan languages
wal		Walamo
war		Waray
was		Washo
wel/cym	су	Welsh
wen		Sorbian languages
wln	wa	Walloon
wol	WO	Wolof
xal		Kalmyk; Oirat
xho	xh	Xhosa
yao		Yao
уар		Yapese
yid	yi	Yiddish
yor	уо	Yoruba
ypk		Yupik languages
zap		Zapotec
zen		Zenaga
zha	za	Zhuang; Chuang
chi/zho	zh	Chinese
znd		Zande languages
zul	zu	Zulu
zun		Zuni
ZXX		No linguistic content
zza		Zaza; Dimili; Dimli; Kirdki; Kirmanjki; Zazaki

Список ссылок на сайты используемых программ

Abbyy Lingvo - http://www.abbyy.ru Abbyy FineReader - http://www.abbyy.ru Kromsator by Bolega http://bolega.hotmail.ru/ EmEditor - http://www.emeditor.com Microsoft Excel http://www.microsoft.com OPФO 9 – http://www.orfo.ru , http://www.informatic.ru OpenOffice Spreadsheets http://ru.openoffice.org , или сайт фирмы Инфра http://i-rs.ru/ (как выразился участник рубордовского форума dg333 – их сборка русее. Hunspell http://hunspell.sourceforge.net Словари к Hunspell и HSpell (потребуют перекодировки в UTF8) <u>ftp://ftp.unimuenster.de/pub/software/OOoDev/dicooofull/pack_full_20070914.zip</u> Hspell - ???

ЧЗВ с форума Ru-Board, форума Lingvoda и мои собственные замечания.

- В: Не удается скомпилировать словарь. Что делать?
- **O:** При компиляции словаря ставим галку напротив «Создавать файл ошибок (*.dde)». После компиляции открываем данный словарь в любом текстовом редакторе и смотрим в чем дело.
- **В:** Не удается скомпилировать словарь. В файле ошибок *.dde написано: Redundant opening tag "[trn]" (Повторный открывающий тег "[trn]").
- O: Такое предупреждение возникает, если при конвертации уже размеченного (т.е. с тэгами) dsl-файла словаря Вы не сняли или поставили галку напротив «Автоматически добавлять разметку, необходимую для поиска по содержимому словаря».

Пояснение: Если эта галка установлена, то при компиляции в начале карточки будет поставлен тег начала зоны перевода [trn], а в конце карточки - тег окончания зоны перевода [/trn]. При этом, если в тексте карточки уже есть тег [trn], поставленный туда создателем словаря, то карточка скомпилирована не будет с соответствующим сообщением об ошибке, т.к. вложение зон индексации одного типа недопустимо по правилам языка DSL.

- B: Не удается скомпилировать словарь. В процессе компиляции выскакивает окошко: Cannot create the dictionary. First card headword not found. (Невозможно создать словарь. Не найден заголовок первой карточки.)
- О: Это происходит, если в словаре есть карточка вида: (слово1)_пробел_ слово2 _пробел_карточка

т.е. среди слов-заголовков, которые относятся к данной карточке, есть слово в скобках, после скобок стоит пробел. В данном случае нужно убрать скобки вокруг слова 1. Или если это слово относится к предыдущей карточке, то переместить его в ту карточку.

- В: Как вставить в теле карточки пустую строку?
- О: Возможные варианты Вариант 1

[m1] \ [/m] Вариант 2 Не проверял но мне кажется и просто: \ (т. е. табулятор + слэш + пробел) будет достаточно Вариант 3 Есть универсальный вариант - строка, содержащая закомментированный пробел: {{ }}

- В: Почему в словаре количество заголовков отличается от количества статей?
- О: Возможные причины

Вариант 1 term1 term 2 term 3 translation <u>Результат компиляции:</u> {{Number of headings(Кол-во заголовков): 3. Number of entries(Кол-во статей): 1.}} Вариант 2 Еще вариант: term(s) text <u>Результат компиляции:</u> {{Number of headings: 2. Number of entries: 1.}}

В: Каков порядок сортировки неалфавитных символов в Лингво

О: Порядок сортировки заголовков в Лингво отличается от принятого в бумажных словарях (где обычно неалфавитные символы вообще не учитываются). Это надо помнить. Я набросал тестовый словарик, чтобы проверить этот и следующий вопрос. Ниже скриншот.



В: О подвешенных ссылках.

О: Посмотрите скриншот выше. Вот оригинал из словаря АСВВА

> [m1][b]1.[/b] [trn]abc1[/trn][/m] [m1][ref]A(CC)BB[/ref][/m] A(CC)BB [m1][b]1.[/b] [trn]abc1[/trn][/m]

Поскольку в бумажных словарях ссылки типа см. A(CC)BB обычное дело, при конвертации в Лингво от них необходимо избавиться. Также нехорошо выглядят и карточки ACCBB и ABB

😑 ACCBB (Pt-Pt)	
<u> Ф</u> ормы слова	🙌 Найти в карточке
A(CC)BB 1. abc1	

В бумажном словаре запись A(CC)BB обычно означает ACCBB и ABB. Поэтому наверное лучше переделывать такие карточки в вариант

ACCBB ABB [m1][b]1.[/b] [trn]abc1[/trn][/m]

- В: Как добавить в словарь пример, не утратив полезную информацию цвет, удароения и т.п.
- Конвертируя один словарь с большим количеством примеров в Лингво я (как и некоторые другие) столкнулся с неприятной особенностью стандартного оформления примеров по методике приведенной в справке к Лингво: Пример

[m2][*][ex][lang id=1033][c blue]to abandon a Bill[/c][/lang] — отказаться от продвижения законопроекта[/ex][/*][/m]

Текст, как оригинала, так и перевода выглядят одинаково (Вы можете увидеть их в файле справки к Лингво),- что не есть хорошо с точки зрения психологии восприятия. Помимо этого теги [*][ех] забивают теги оформления – ударение, тип шрифта.

Для улучшения восприятия можно попытаться заменить эти тэги на [lang} и [c], как показано на примере ниже.

[m2][lang id=1033][c blue]to abandon a Bill[/c][/lang] — [lang id=2]отказаться от продвижения законопроекта[/lang][/m]

И вот так это выглядит на практике

[m2][lang id=5][c blueviolet]a nov[']e[/']mbre[/c][/lang] — [lang id=2]в ноябре[/lang][/m] а novembre — в ноябре

Отрицательным моментом данного решения является невожможность поиска по примерам. Ну тут уж кому и что важнее.

В: Где посмотреть, как выглядят используемые в Лингво цвета шрифтов.

О: Ответ на этот вопрос есть в справке к Лингво (как в Интернет Explorer 4.0). Жаль только не приведена гамма, где каждый цвет указан и подписан. Существует маленький словарь colors8.lsd, в котором можно просмотреть все(?) 139 используемых цветов. Чтобы избавить Вас от необходимости его поисков и подключения приведу здесь полный скриншот. Он даст более или менее точное представление того, что вы получите при использовании выбранного цвета.

Color samples

aliceblue: antiquewhite: Texcr. напечатанный userow antiquewhite aqua: Текст, напечатанный цветом aqua aquamarine: Текст, напечатанный цветом aquamarine azure: beige: Текст, калематалнык изеток beige bisque: Текст, напечатанный цеетом bisque blanchedalmond: Tekcri напечатанный userow blanchedalmond blue: Текст, напечатанный цветом blue blueviolet: Текст, напечатанный цветом blueviolet brown: Текст, напечатанный цветом brown burlywood: Текст, напечатанный цветом burlywood cadetblued: Текст, напечатанный цветом cadetblue chartreuse: Текст, напечатанный цветом chartreuse chocolate: Текст, напечатанный цветом chocolate coral: Текст, напечатанный цветом coral cornflower: Текст, напечатанный цветом cornflower cornsilk: crimson: Текст, напечатанный цветом crimson cyan: Текст, напечатанный цветом суап darkblue: Текст, напечатанный цветом darkblue darkcyan: Текст, напечатанный цветом darkcyan darkgoldenrod: Текст, напечатанный цветом darkgoldenrod

darkgray: Текст, напечатанный цветом darkgray darkgreen: Текст, напечатанный цветом darkgreen darkkhaki: Текст, напечатанный цветом darkkhaki darkmagenta: Текст, напечатанный цветом darkmagenta darkolivegreen: Текст, напечатанный цветом darkolivegreen darkorange: Текст, напечатанный цветом darkorange darkorchid: Текст, напечатанный цветом darkorchid darkred: Текст, напечатанный цветом darkred darksalmon: Текст, напечатанный цветом darksalmon darkseagreen: Текст, напечатанный цветом darkseagreen darkslateblue: Текст, напечатанный цветом darkslateblue darkslategray: Текст, напечатанный цветом darkslategray darkturquoise: Текст, напечатанный цветом darkturquoise darkviolet: Текст, напечатанный цветом darkviolet deeppink: Текст, напечатанный цветом deeppink deepskyblue: Текст, напечатанный цветом deepskyblue dimgray: Текст, напечатанный цветом dimgray dodgerblue: Текст, напечатанный цветом dodgerblue firebrick: Текст, напечатанный цветом firebrick floralwhite: forestgreen: Текст, напечатанный цветом forestgreen fuchsia: Текст, напечатанный цветом fuchsia gainsboro: Текст, напечатанный цветом gainsboro ghostwhite: gold: Текст, напечатанный цветом gold goldenrod: Текст, напечатанный цветом goldenrod gray: Текст, напечатанный цветом gray green: Текст, напечатанный цветом green greenyellow: Текст, напечатанный цветом greenyellow honeydew: hotpink: Текст, напечатанный цветом hotpink indianred: Текст, напечатанный цветом indianred indigo: Текст, напечатанный цветом indigo ivory: khaki: Текст, напечатанный цветом khaki lavender: Texct, напечатанный цветом lavender lavenderblush: lawngreen: Текст, напечатанный цветом lawngreen lemonchiffon: The submersion and an and a lemon billion lightblue: Текст, напечатанный иветом lightblue lightcoral: Текст, напечатанный цветом lightcoral lightcyan: Terror is the statistic been of induced an lightgoldenrodyellow: Texts management and a service lightgreen: Текст, напечатанный цветом lightgreen lightgray: Текст, напечатанный цветом lightgray lightpink: Текст, напечатанный цветом lightpink lightsalmon: Текст, напечатанный цветом lightsalmon lightseagreen: Текст, напечатанный цветом lightseagreen

lightseagreen: Текст, напечатанный цветом lightseagreen lightskyblue: Текст, напечатанный цветом lightskyblue lightslategray: Текст, напечатанный цветом lightslategray lightsteelblue: Текст, напечатанный цветом lightsteelblue lightyellow: lime: Текст, напечатанный цветом lime limegreen: Текст, напечатанный цветом limegreen linen: Texco kanesaraksuk paaroo liner magenta: Текст, напечатанный цветом magenta maroon: Текст, напечатанный цветом maroon mediumaquamarine: Текст, напечатанный цветом mediumaguamarine mediumblue: Текст, напечатанный цветом mediumblue mediumorchid: Текст, напечатанный цветом mediumorchid mediumpurple: Текст, напечатанный цветом mediumpurple mediumseagreen: Текст, напечатанный цветом mediumseagreen mediumslateblue: Текст, напечатанный цветом mediumslateblue mediumspringgreen: Текст, напечатанный цветом mediumspringgreen mediumturquoise: Текст, напечатанный цветом mediumturquoise mediumvioletred: Текст, напечатанный цветом mediumvioletred midnightblue: Текст, напечатанный цветом midnightblue mintcream: mistyrose: moccasin: Текст, напечатанный цеетом moccasin mediumorchid: Текст, напечатанный цветом mediumorchid mediumpurple: Текст, напечатанный цветом mediumpurple mediumseagreen: Текст, напечатанный цветом mediumseagreen mediumslateblue: Текст, напечатанный цветом mediumslateblue mediumspringgreen: Текст, напечатанный цветом mediumspringgreen mediumturquoise: Текст, напечатанный цветом mediumturquoise mediumvioletred: Текст, напечатанный цветом mediumvioletred midnightblue: Текст, напечатанный цветом midnightblue mintcream: mistyrose: Terct, капечетанный userow mistyrose moccasin: Текст, напечатанный цветом moccasin navaiowhite: Текст, напечатанный изэтся navaiowhite navy: Текст, напечатанный цветом navy oldlace: olive: Текст, напечатанный цветом olive olivedrab: Текст, напечатанный цветом olivedrab orange: Текст, напечатанный цветом orange orangered: Текст, напечатанный цветом orangered orchid: Текст, напечатанный цветом orchid palegoldenrod: Текст, напечатанный уветом palegoldenrod palegreen: Текст, напечатанный цветом palegreen paleturquoise: Текст, напечатанный цветом paleturquoise palevioletred: Текст, напечатанный цветом palevioletred papayawhip: Terrer Haneyaralekshi userow papayawhip

peachpuff: Teкct, напечатанный иветом peachpuff peru: Текст, напечатанный цветом peru pink: Текст, напечатанный цветом pink plum: Текст, налечатанный цветом plum powderblue: Текст, напечатанный цветом powderblue purple: Текст, напечатанный цветом purple red: Текст, напечатанный цветом red rosybrown: Текст, напечатанный цветом rosybrown royalblue: Текст, напечатанный цветом royalblue saddlebrown: Текст, напечатанный цветом saddlebrown salmon: Текст, напечатанный цветом salmon sandybrown: Текст, напечатанный цветом sandybrown seagreen: Текст, напечатанный цветом seagreen seashell: sienna: Текст, напечатанный цветом sienna silver: Текст, напечатанный цветом silver skyblue: Текст, напечатанный цветом skyblue slateblue: Текст, напечатанный цветом slateblue slategray: Текст, напечатанный цветом slategray snow: springgreen: Текст, напечатанный цветом springgreen steelblue: Текст, напечатанный цветом steelblue tan: Текст, напечатанный цветом tan teal: Текст, напечатанный цветом teal thistle: Текст, напечатанный цветом thistle tomato: Текст, напечатанный цветом tomato turquoise: Текст, напечатанный цветом turquoise violet: Текст, напечатанный цветом violet wheat: Текст, напечатанный цветом wheat white: whitesmoke: yellow: Текст, напечатанный цвэтом yellow yellowgreen: Текст, напечатанный цветом yellowgreen (Всего 139 цветов.)

- В: При подключении к Лингво одноязычный словарь добавляется к двуязычным направлениям (например, русско-русский словарь попадает еще и на русско-
- английские, русско-немецкие и т.д. полки) О: НЕЙТРАЛИЗАЦИЯ ПОДКЛЮЧЕНИЯ ОДНОЯЗЫЧНЫХ ПОЛЬЗОВАТЕЛЬСКИХ СЛОВАРЕЙ К ДВУЯЗЫЧНЫМ НАПРАВЛЕНИЯМ (С форума Ru-board).

Начиная с 10-й версии, ABBYY Lingvo принудительно подключает все одноязычные словари к двуязычным направлениям, у которых первый язык совпадает с одноязычным словарем. Наибольшие проблемы возникают с русско-русскими словарями, во-первых, потому что их много, во-вторых, потому что в ABBYY Lingvo 11ML по дефолту девять русско-нерусских направлений, а у тех, кто подключал дополнительные языки и того больше. Теоретически можно отключить одноязычные словари, просто выключив после их установки соответствующие чек-боксы, но если вы подключили 80 русско-русских словарей, что далеко не предел, то таких чек-боксов будет минимум 720! Необходимо также учесть, что Lingvo перемешивает словари двух направлений между собой, не давая возможности отсортировать их по дате или направлению, так что вам придется не просто механически отключать чек-боксы, а внимательно вычитывать минимум девять длинных списков, стараясь не удалить лишнее. Добавьте к этому, что многие пользовательские

словари не имеют суффикса языковой пары (Ru-Ru, Ru-En,...), а определить только по имени к какому направлению относится словарь не всегда возможно. А если у вас два компьютера - дома и на работе? Короче работа долгая, тупая и неблагодарная на которую я, например, оказался не способен. Вот и пришлось придумывать методы обхода этого бага.

ВАРИАНТ 1 - "ОТВЛЕЧЕНИЕ НА НЕГОДНЫЙ ОБЪЕКТ" Вариант 1 очень прост, но ограничен только вновь подключаемыми словарями.

Обратите внимание, что Lingvo 11 подключает одноязычные словари не к двуязычной полке словарей в целом, а к конкретной группе этой полки, той, которая активна во время установки одноязычного словаря (обычно это самая важная "Основная группа"). То, что само двуязычное направление не активно и даже отключено значения не имеет.

РЕШЕНИЕ ПРОБЛЕМЫ:

Создаем в наиболее уязвимых языковых парах, прежде всего русско-нерусских и английско-неанглийских группу обманку. Я назвал ее _Null - символ подчеркивания делает ее более заметной в списке групп, но вообще имя группы-обманки может быть любым. Теперь перед подключением одноязычных словарей достаточно сделать активной _Null группу в двуязычных направлениях, у которых первый язык совпадает с одноязычным словарем. Конечно это вариант тоже слегка нагрузочен, но, во-первых, переключить 9 групп легче, чем выискивать и отключать словари в 9 длинных смешанных списках, а во-вторых, большинство редко использует все 9 русско-нерусских направлений, так что фактически переключать придется 3-4 группы, в остальных направлениях _Null-группа будет оставаться активной практически постоянно.

Ну а что делать, если вы уже успели подключить сотню - другую одноязычных словарей и теперь, при мысли о необходимости их выковыривания из как минимум 9-ти направлений, у вас начинается головная боль? Чтобы облегчить себе жизнь придется немного поработать с реестром.

Специально для тех, кто знает о peecтpe Windows только то, что туда лучше не лазить, дается максимально подробное описание работы. Из-за этого получилось несколько длинно, занудно, но зато безопасно. На практике же все быстро, просто и занимает от силы 2-3 минуты.

ВАРИАНТ 2 - НЕЙТРАЛИЗАЦИЯ ПОДКЛЮЧЕНИЯ ОДНОЯЗЫЧНЫХ СЛОВАРЕЙ К ДВУЯЗЫЧНЫМ НАПРАВЛЕНИЯМ ПРИ ПОМОЩИ МАНИПУЛЯЦИЙ С РЕЕСТРОМ.

У тех, кто инсталлировал Lingvo 11 по дефолту, конфигурация полок и словарей хранится отдельно для каждого пользователя в разделе реестра HKEY_CURRENT_USER: HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries

У тех, кто при инсталляции выбрал "общие настройки для всех пользователей" конфигурация полок и словарей хранится в разделе реестра HKEY_LOCAL_MACHINE:

HKEY_LOCAL_MACHINE\SOFTWARE\ABBYY\Lingvo\11.0\Multilingual\Dictionaries

Если вы не помните как настроена ваша Lingvo, зайдите в любой из указанных разделов реестра и посмотрите в каком из них находится ветка Configuration

в

HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configuration или в

HKEY_LOCAL_MACHINE\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configuration

Если в HKEY_CURRENT_USER\... значит конфигурация полок и словарей хранится отдельно для каждого пользователя, ну а если в HKEY_LOCAL_MACHINE\... то конфигурация для всех общая.

ВАРИАНТ 2.1

1. Заходим в реестр [Пуск - Выполнить - regedit - OK] или [Start - Run - regedit - OK] в ветку HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries

или HKEY_LOCAL_MACHINE\SOFTWARE\ABBYY\Lingvo\11.0\Multilingual\Dictionaries (для пользователей 11SL меняем здесь и далее Multilingual на SixLanguages) и экспортируем раздел Configuration [правый клик на имени раздела - "экспортировать" дать файлу осмысленное имя, например: Configuration_0].

2. Устанавливаем в обычном порядке одноязычные словари на нужные полки (Ru-Ru, En-En, It-It и т.д.). Двуязычные словари желательно устанавливать либо до экспорта раздела Configuration, либо после окончания манипуляций с одноязычными словарями.

3. После установки всех одноязычных словарей закрываем окно Lingvo, щелкаем правой клавишей на иконке Lingvo в трее и закрываем программу - "Завершить работу Lingvo". 4. Открываем раздел реестра

HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configuration или

HKEY_LOCAL_MACHINE\SOFTWARE\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configurat ion и экспортируем разделы, в которые были установлены одноязычные словари. Коды языков приведены в Справке Lingvo: "Справка – Приложение - DSL Compiler -Поддерживаемые языки", но для удобства приведу основные 10 кодов:

1028 - Ch, 1033 - En, 1034 - Sp (SpanishTraditionalSort), 1036 - Fr, 1040 - It, 1049 - Ru, 1055 - Tu, 1058 - Uk, 1540 - La, 32775 - Ge(GermanNewSpelling).

Следовательно, если мы подключали словари Ru-Ru, En-En и It-It, нам надо экспортировать разделы: 1049-1049,1033-1033 и 1040-1040

5. Теперь удаляем раздел реестра Configuration (HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configuration или

HKEY_LOCAL_MACHINE\SOFTWARE\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configurat ion)

6. Импортируем раздел Configuration сохраненный перед установкой одноязычных словарей [Щелкаем два раза по ранее экспортированному файлу Configuration_0].

7. Удаляем в нем разделы соответствующие установленным перед этим одноязычным словарям (в нашем случае 1049-1049,1033-1033 и 1040-1040) после чего импортируем сохраненные после! установки одноязычных словарей разделы (в нашем случае 1049-1049,1033-1033 и 1040-1040) - теперь все чисто, включаем Lingvo.

А как быть с путями к файлам словарей? А никак! Они прописаны в других ветках реестра и мы их пока не трогали!

ВАРИАНТ 2.2

ОЧИСТКА LINGVO ОТ ОДНОЯЗЫЧНЫХ СЛОВАРЕЙ ПОДКЛЮЧЕННЫХ К ДВУЯЗЫЧНЫМ НАПРАВЛЕНИЯМ РАНЕЕ.

Если вы сначала установили одноязычные словари и только потом поняли, какой кошмар вас ожидает, восстановление некоторых разделов реестра облегчит вам жизнь.

Для восстановления реестра вам понадобятся файлы подразделов ветки Configuration, сделанные сразу после инсталляции. У вас их, возможно, нет, но, поскольку этот раздел ПК-независим, я выложил их в своей папке у KVK (только для Lingvo 11ML).

Предположим, вы уже установили 80 русско-русских словарей, и они замусорили все 9 (а если установлены дополнительные языки то и больше) русско-нерусских направлений и перспектива отключать минимум 720 словарей из смешанных списков вам не улыбается.

Предположим также, что раздел Ru-En вы все-таки почистили и установили туда дополнительные словари, а в раздел Ru-Sp несколько дополнительных словарей уже установили, но от 80-ти словарей Ru-Ru еще не очищали.

ЧИСТИМ!

1. Прежде всего, на всякий случай, делаем копию (экспорт) раздела HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configuration или

HKEY_LOCAL_MACHINE\SOFTWARE\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configurat ion

Для чистки она вам не понадобится, но теперь, чтобы вы не напутали, можно вернуть реестр к рабочему состоянию.

2. Закрываем окно Lingvo, щелкаем правой клавишей на иконке Lingvo в системном трее и закрываем программу - "Завершить работу Lingvo".

3. Открываем раздел

HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configuration или

HKEY_LOCAL_MACHINE\SOFTWARE\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configurat ion

и удаляем все русско-нерусские разделы 1049-хххх, кроме, в нашем примере, разделов Ru-En (1049-1033) и Ru-Sp(1049-1034). Раздел Ru-Ru (1049-1049) естественно тоже не трогаем.

4. Запускаем соответствующие удаленным веткам файлы реестра 1049-хххх.reg сделанные сразу после инсталляции - в них только системные словари, (1049-1049.reg я, чтобы по инерции не нажали, переименовал в _1049-1049.reg).

5. Теперь, а лучше раньше, надо решить, что делать с разделом Ru-Sp (1049-1034)? Если количество установленных дополнительных русско-испанских словарей явно больше чем русско-русских и полок много понаделали, можно оставить пару Ru-Sp, как есть и вычистить русско-русские словари вручную через оболочку Lingvo. Если же ситуация обратная, стоит поступить с направлением Ru-Sp(1049-1034) как с другими парами 1049хххх, а после, в оболочке Lingvo, нажать стрелочку справа от словарей Ru-Sp и подключить нужные. Переустанавливать словари не надо, они были деактивированы для указанных направлений, но не удалены!

Аналогично можно поступить во время или после установки других одноязычных словарей, например En-En.

В: Как быстро перенести все словари Лингво на другой компьютер

O: Ответ пользователя Umid Kulijanov на форуме Ru-Board для Лингво 11 Многоязычный.

Сохранение Конфигурации

- 1. Установить Лингво.
- 2. Хранить ВСЕ нужные пользовательские словари в D:\Lingvo_dics (удобно при восстановлении/переустановке системы). По идее они уже разложены по языковым парам, если нет, то это уже на личный вкус. Создать En-Ru и Ru-En, и др. необходимые языковые пары. Расположить словари по соответствующим папкам.
- 3. Добавить ВСЕ нужные словари в необходимые языковые пары, выполнить индексацию (индексацию проводить после каждого большого словаря или нескольких средних. Выполнять индексацию после добавления ВСЕХ словарей терпения не хватит [©]).

- 4. Закрыть Лингво и выгрузить агента из системного трея (при условии, что Лингво был активен).
- 5. HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\User сохранить ветку. (В ветке расположена информация о расположении/принадлежности к конкретной языковой группе пользовательских словарей, а также полный путь к словарю, т.е. Словарь номер1 --> D:\Lingvo_dics\En-En\Abbrev\abbrev.lsd, и т.д.)
- 6. HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configu ration\1033-1049 сохранить ветку En-Ru (в принципе нужен только Профиль4 т.е. дефолтная группа). Здесь информация о том, какие словари расположены на данной полке.
- 7. HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configu ration\1049-1033 сохранить ветку Ru-En (в принципе нужен только Профиль4 т.е. дефолтная группа). Здесь информация о том, какие словари расположены на данной полке.
- 8. Повторить пункты 6-7 для других необходимых языковых пар (русс.-немец., немец.русс., и т.д. Цифровое обозначение языковой группы см. в Справке к Лингво).
- 9. Все необходимые файлы готовы.

Установка на другие компьютеры / Переустановка системы

- 1. Копируем папку D:\Lingvo_dics на необходимые компьютеры. В случае переустановки системы данный пункт пропустить.
- 2. Установить Лингво.
- 3. Закрыть Лингво и выгрузить агента из системного трея (при условии, что Лингво был активен).
- Импортируем в реестр сохранённую ветку HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\User – user.reg
- 5. Импортируем в реестр сохранённую ветку HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configu ration\1033-1049 – En-Ru.reg
- Импортируем в реестр сохранённую ветку HKEY_CURRENT_USER\Software\ABBYY\Lingvo\11.0\Multilingual\Dictionaries\Configu ration\1049-1033 – Ru-En.reg
- 7. Повторить пункты 5-6 для других необходимых языковых пар, ветки реестра которых былы сохранены (русс.-немец., немец.-русс., и т.д.
- 8. Заходим в Tools → Dictionaries. Нажать Update Index...
- 9. Лингво 11 должен сообщить что «Все словари уже проиндексированы», правда, после продолжительного «раздумья». Т.к. для каждого словаря создаётся отдельный индекс в папке с данным словарём.

Данная процедура должна подходить (теоретически) и для других версий Лингво, необходимо будет только изменить путь реестра с учётом установленной версии. И после выполнения всех вышеперечисленных шагов набраться терпения в ожидании завершения индексации. Также как вариант, импортировать и индексировать только по одной языковой паре за раз.

Создание и подключение иконки к словарю в Лингво.

Поскольку не все конвертеры словарей в Лингво художники для них остаётся два варианта получения иконки – попросить (заказать) ее у специалиста, или прямиком на <u>http://www.google.com</u>. Гугль имеет одну полезную фичу, которая нам поможет - поиск картинок



Итак, жмем Картинки и вводим (желательно на английском языке) что-то, что однозначно ассоциируется у нас с тематикой словаря – например для ядерного словаря это может быть «atomic blast» или «radioactivity danger sign» и выбираем для начала показывать только маленькие картинки.



Ну вот, будем считать, нам повезло, и вторая картинка пойдет для нашей иконки. Кликаем по ней мышой и выбираем Сохранить рисунок как (Это в Опере. Как в Эксплорере или Огненной лисе не знаю – Вы сами должны знать) и Сохраняем рисунок. Далее в любимом

графическом редакторе доводим его до размера 14Х21 пиксель 🍒 и сохраняем в bmpформате с названием таким, как у нашего словаря.

Копируем иконку в каталог со словарем и компилируем его Если иконка правильная, то при компиляции она отобразится справа от надписи «Значок» компилятора

🎯 DSL Compiler	? ×
Свойства исходного файла	· · · ·
Файл: C:\Downloads\nuc.dsl	💌 Выбрать
Значок: 🔶	

После компиляции подключаем откомпилированный словарь и индесируем его. Если иконка не появилась на панели в Лингво - выгружаем Лингво и загружаем его снова. Вот вроде и все и напоследок несколько советов.

- Поскольку размер иконки в 12 Лингве 14х21 старайтесь выбирать изображения прямоугольные, с пропорцией 2:3 (ширина к высоте). Посколько иконки и рисунки обычно квадратные или имеют бо́льшую ширину, то смотрите на картинки с большими боковыми полями, которые не жалко будет обрезать.
- 2. Картинки обрезайте в пропорции 2:3. (20х30, 40х60 и т.п.)
- 3. Если не нашли пригодную маленькую картинку, переключайтесь на больший размер.

В целях упорядочения иконотворчества хотелось бы, чтобы в этом вопросе была какая-то стандартизация. Поэтому ниже привожу список стандартных иконок из Лингво плюс некоторые другие понравившиеся мне иконки.

Словарь		
Универсальный		
Collins	0	
Неформальный		
Идиомы	Ģ	
Американа		
Англия	20 20 20	
Грамматический		
Компьютеры	S	
Компьютеры от Лингво	c 📕	
Телеком		
Маркетинг	_	
Финансовые рынки	2	
Менеджмент	8	
Финансовый менеджмент		
Бухучёт		
Экономика	0	
Юридический	4	
Патенты	6	
Политехнический	Ā	
Физика	U	
Медицина	¥	
Биология	2	
Инжиниринг	0	
Химия	4	
Строительство	1	
Автомобили		
Пищевая промышленность, еда	21	
Машиностроение и автоматизация производства	10	
Нефть и газ	*	
Научно-технический		
Вино		
Пиво	9	
Искусство	â	
Автосервис		
ХИМИЯ И ПЕРЕРАБОТКА НЕФТИ	N	
МАШИНОСТРОЕНИЕ И МЕТАЛЛООБРАБОТКА	2	
СВАРКА	2	
СЕЛЬСКОЕ ХОЗЯЙСТВО	1 🏠 🏠 🕅	
ПАРФЮМЕРИЯ И КОСМЕТИКА		
ДИПЛОМАТИЯ	V	
	-	

ЭЛЕКТРОНИКА	۲
ТЕКСТИЛЬ	2
ПОЛИМЕРЫ, БИОХИМИЯ	S
МЕТАЛЛУРГИЯ	3
ЭКОЛОГИЯ	(9
КОНТРОЛЬ КАЧЕСТВА	۲
ЯДЕРНЫЙ	2
ТЕПЛОТЕХНИКА	1
МЕТРОЛОГИЯ	T _a
ЖЕЛЕЗНАЯ ДОРОГА	臭

Работа с Hspell.

Внимание!!!

1) Сохраняем отформатированный в Emeditor файл в кодировке Utf-8 с BOM

<u>И</u> мя файла:	Untitled	•	Со <u>х</u> ранить
<u>Т</u> ип файла:	Text (*.bxt)	•	Отмена
Encoding:	Unicode 💌]	
	Add a Unicode Signature (BOM)		
Return Method:	No Change	[

и в последующем, если что-то не работает первым делом проверяем везде кодировку.

Step1

Получение списка всех слов из указанного текста

Щелкнем правой кнопкой мыши (ПКМ) в среднем окошке... и в контекстном меню выбираем пункт Get list of all words from text.

E HSpell 1.4.3	
Help Main □ it_IT_correct □ ru-ru_stems □ ru_RU □ ru_RU_ie □ it_IT_correct spelled from text Remove correct spelled ford Ignore case Save as If Customize Font	S
0 0	
	Select list
	Select text Start
0	

В открывшемся окне выбираем наш файл в кодировке Utf-8 с BOM
Открыть				<u>? ×</u>
Папка:	Camples	•	-]▼
Недавние документы Рабочий стол Мои документы Мой компьютер	schkarov0.txt	UTF-8		
Сетевое окружение	Имя файла:	schkarov0.txt	-	Открыть
alos"	Тип файлов:	Text file to process (*.txt)		Отмена
		🗖 Только чтение		k

Программа автоматически создаст список слов, который Вы видите в среднем поле.

🗮 HSpell 1.4.3		
Help		
Main		
I IT_correct I IT_cru_stems I I IT_ru_RU I I IT_ru_RU I I IT_cru_RU I I IT_cru_RU I I IT_cru_RU_ie I I I I I I I I I I I I I I I I I I I	\$тренней *вагон *y /3 /a /accettante /aerea /e /p/ /per /pi /Pi /eц /ж /Й /Ка /O	
0 0		
	/	Select list
		Select text Start
	27998	

Список не содержит дубликатов и отсортирован



Удаление из полученного списка правильно написанных слов

В первом поле находятся словари. Это словари из Open Office (пара файлов xx.aff и xx.dic) но в кодировке <u>UTF8!!!</u> (Стандартно ru_RU_уо словарь имеет кодировку KOI).

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

Выбираем необходимые словари, расставляя флажки.

HSpell 1.4.3		
Help		
Main		
IT_correct	\$тренней	
✓ □ ru-ru_stems	*вагон	
	/3	
	/a	
	/aerea	
	/e	
	/p/ /per	
	/pi	
	/PI /eu	
	/ж	
	/// //////////////////////////////////	
	/o	
	/0	
0 0		
		Select list
		Select text Start
 	27998	

Затем щелкнаем правой кнопкой мыши (ПКМ) в среднем окошке... и в контекстном меню выбираем пункт «**Remove correct spelled words**».

E HSpell 1.4.3	
Help	
Main	
II_ it_IT_correct II ru-ru_stems II ru_RU II ru_RU_ie II ru_RU_yo	\$тренней *Baroн *V /3 /a /accettante /aerea /e Get list of all words from text /p/ Remove correct spelled words /per Ignore case /pi Save as /Pi Save as /Pi Save as //i //i //i //i //i //i //i
0 0	
	Select list
	Select text Start
	27998

Получаем результат в строке состояния

HSpell 1.4.3		
Help		
Main		
it_IT_correct □ ru-ru_stems □ ru_RU □ ru_RU_ie □ ru_RU_yo	\$тренней *Вагон 'Y /3 /accettante /acrea /e /pi /pi /pi /Pi /eц /ж /й /acrea /e /pi /pi /Pi /eц /% /й /acrea /pi /pi //Acrea /acrea /acrea /bi /bi /acrea /acrea /acrea /acrea /acrea /bi /bi <	it_IT_correct : passed ru-ru_stems : passed ru_RU : passed ru_RU := : passed ru_RU_e : passed ru_RU_yo : passed
0 0		
		Select list
		Select text Start
	3605	

Если установить флажок «Ignore case»

	1.0	antove contract apareat worka
	Ig	nore case
 /	'i	
/e	щ	✓ Ignore case
/x	ĸ	Save as
11	1	Tr Customize Font
/k	a	

то перед проверкой все символы в словах будут переводиться в нижний регистр

Step3: Сохранение списка в файл

HSpell 1.4.3		
Help		
Main		
☐ ît_IT_correct ☐ îru-ru_stems ☐ îru_RU ☐ îru_RU_ie ☐ îru_RU_yo	\$тренней it_IT_correct : passed *вагон ru-ru_stems: passed *v ru-RU : passed /a ru-RU : passed /a ru-RU : passed /accettante ru-RU : passed /accettante ru-RU : passed /acrea /e /pi Get list of all words from text /pi Remove correct spelled words /k ✓ Ignore case /ň Save as /o /a /a ✓	
0 0		
	Select list	
	Select text Start	
	3605	

Имеется выбор, сохранять полученный список в одноколоночный файл или в двухколоночный (имеет расширение hsp), для последующего исправления и автоматической замены.



Формат списка с двумя колонками:

^слово\t\|\tслово\$

Step 4: Собственно исправление ошибок

Двухколоночный список впоследствии можно использовать для автоматической замены неправильных слов (слева) на исправленные (справа).

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

Поэтому исправлять слова надо в правой колонке Если удалить правую колонку, то при замене слово будет удалено Элементы списка не отвечающие условию ^([^\t\x20]+)\t\|\t([^\t]*)\$ будут проигнорированы программой

Пример:		
до		
ЮрИСДЙКЦИ/я		ЮрИСДЙКЦИ/я
после		
ЮрИСДЙКЦИ/я	I	юрисдикция

При замене учитывается регистр слов в левой колонке

Step 5: Соственно , исправление ошибок

Нажимаем кнопку _____ и выбираем сохранённый файл с исправлениями



Затем, нажимаем кнопку Select text... и выбираем отформатированный файл словаря (в UTF 8)

Открыть				? ×	HSpell Beta 1.3			_ 🗆 🛛
Папка:	C Sample		-		Main			
Недавние документы Рабочий стол Мои документы Документы Мой	tst schkarov0.txt				it_IT ☐ ru-ru_stems ☐ ru_RU ☐ ru_RU_ie ☐ ru_RU_jo	\$тренней "вагон "y /3 /accettante /aerea /e /p/ /pr /pr /Pi /Pi /ж /й	it_IT : pas: u-ru_stem ru_RU : pa ru_RU_po Current list 3851 Get lis Remov	sed s: passed : passed : passed was saved Save t of all words from text e correct spelled words
компьютер Сетевое окружение	Имя файла: Г Тип файлов: Г	schkarov0.txt Text file in UTF-8 encodin;	g to process (".txt)	Открыть Отмена	Accepted: 3851 Ignor [C:_24\HSpell_Beta_1_; [C:_24\HSpell_Beta_1_;	red: 0 3\Sample\mylist.list 3\Sample\schkarov0.txt		Select list
					<u></u>			
Нажима	ем кнопн	Start						

После этого сохраняем результат.

Сохранить как							2 ×
<u>П</u> апка:	🔁 Sample			•	🗕 🔁 🚔 🖩		
Недавние документы Рабочий стол Мои	t <u>×t</u> lschkarov0.txt						
документы Мой компьютер Сетевое окружение	<u>И</u> мя файла: <u>Т</u> ип файла:	result.txt Text file (*.txt	:)		•	Сохрани Отмен	ть

Внимание!!! Если вы введете имя уже существующего файла, он будет перезаписан

Пополнение словарного запаса программы HSPELL

Щёлкаем ПКМ в левом окошке и выбираем команду контекстного меню «Create user dictionary»

🗮 HSpell 1.4.3		
Help		
Main		
it_IT_correct it_IT_correct it_IT_correct ru_ru_stems ru_RU ru_RU_ie it_RU_ie it_RU_yo Create user dictional Update	Stренней *Baroн *V /3 /a /accettante /aerea /e /p/ /per V····	it_IT_correct : passed ru-ru_stems : passed ru_RU : passed ru_RU_ie : passed ru_RU_yo : passed Current list was saved Current list was saved
	/Й /ка /о /О /я Гавто]	
0 0		
		Select list
		Select text Start
1		
	3605	

В появившемся окне выбираем язык

// Dictionary		
	#French user dictionary SET UTF-8 TRY eiarsntolucéÉmpdgbhfvqjyxèĚzâÅkêÊçÇîÎlīwûÜôÔàÀeËüÜöööóÓĺİùÙáÁa	French
French		French English German Italian Russian Spanish
C Files	Start	Ukrainian

В поле ниже вводим имя файла и нажимаем кнопку «Submit». При этом кнопка «Start» активируется.

Указываем местонахождение списка слов («Clipboard» (Буфер обмена) или «Files» (Файлы)) (Можно выбрать сразу несколько файлов)



Словарь делается из одноколоночных списков слов (одна колонка). Содержимое файлов со списками полностью загружается в оперативную память.

В. Можно ли туда например загрузить гигабайтный файл?

О. Не пробовал. Лучше наверно сделать несколько словарей

Для добавления нового словаря без перегрузки программы используйте команду контекстного меню «Update»

E HSpell 1.4.3	
Help	
Main	
Create user dictionary Update	
0 0	
	Select list
	Select text Start
0	

Это будущее содержание *.aff



Словари создаются в папке dict

Помните!!!

- 1) Вычитываемый текст словаря должен иметь кодировку Utf-8 и только Utf-8.
- 2) СПИСКИ для создания нового словаря могут иметь либо Utf-8, либо Utf-16LE (Unicode)

ОБЯЗАТЕЛЬНО с ВОМ

3) Если что-то не работает, то первым делом проверяем везде кодировку.

Меню Setup

Здесь можно настроить разделители слов и знаки препинания, по которым текст будет делиться на слова и которые будут удаляться из списка.

🚰 Setup	_ 🗆 ×
Main	
Dividere	
$ - r^{n} \cdot i 0 \sim i \cdot m -$	
Submit	

Введите символы в нижнее (белое) поле и нажмите кнопку «Submit»

С этого момента программа использует новый список разделителей

Стандартно используются следующие разделители

🔀 Setup	<u>_ 🗆 ×</u>
Main	
Dividers: ~ \".,;+!{}"<>«»?: Default ~ \.,;+!{}"<>?: «»-s Submit	

Слэши (\) перед {}[]()^\$.|*+?\ ставить не надо

Это не регулярное выражение, а просто перечень символов разделителей

Итог кое-каких опытов с hunspell'овскими словарями

а)Сортировка слов в словаре не влияет на его свойства

(т.е. работа с ним в обоих случаях принесет те же результаты)

mydic.dic		mydic.dic	
zufolii	=	abbeverato	
orfanella		orfanella	
abbeverato		zufolii	

б)Наличие двойников в словаре не влияет на его свойства

(т.е. работа с ним в обоих случаях принесет те же результаты)

Автор: <u>niccolo</u> + andreyefg + ru-board Версия: 3.0

zufolii = abbeverato

orfanella orfanella

abbeverato zufolii

zufolii

zufolii

в) Регистр

* Если все символы слова в нижнем регистре,

то результат будет таким:

яшма

input output яшма да Яшма да ЯШМА да яШмА нет

* Если у слова только первый символ в верхнем регистре, то результат будет таким:

Арсеньевич

input output

Арсеньевич да

арсеньевич нет

АРСЕНЬЕВИЧ да

АрсеНьевИч нет

* Если все символы слова в верхнем регистре, к примеру АЗС,

то результат будет таким:

A3C

input	output
A3C	да
азс	нет
Азс	нет
азС	нет