

# Преобразование словаря из формата Excel в формат Abbyu Lingvo

В данном документе описан алгоритм преобразования словарей из MS Excel в формат словарей Abbyu Lingvo. Алгоритм пригоден для любых словарей, составленных в виде таблицы, которую можно скопировать в MS Excel или сохранить в виде текстового файла со знаками табуляции между колонками. Преобразование выполняет программа<sup>1</sup>, написанная мной на языке Python. Программа сортирует записи по алфавиту, устраняет некоторые ошибки в оформлении, формирует словарные статьи, добавляет разметку языка DSL (язык Лингво для составления словарей) и выдает файл в формате DSL. Пока не реализовано преобразование кодировки этого файла в Unicode. Это надо делать вручную. Моя основная специализация – технический перевод (промышленность, строительство, нефть и газ и т.п.). В программировании опыт небольшой, можно было сделать лучше, тем не менее всё работает, как задумано.

Более подробно причинах выбора форматов и подходе к составлению глоссариев<sup>2</sup> изложен в конце документа в разделе «Обоснование». Ниже инструкция по преобразованию на примере англо-русского глоссария.

## Подготовка

Исходный глоссарий должен быть составлен в формате MS Excel с отдельными листами для направлений английский – русский, русский – английский и (если есть глоссарии без перевода, только с объяснением значений) русский – русский, английский – английский.

В двуязычном словаре должно быть ровно 6 колонок:

1. Термин.
2. Сокращенный или альтернативный вариант термина.
3. Перевод.
4. Сокращенный или альтернативный вариант термина на втором языке.
5. Определение, объяснение, примеры употребления.
6. Источник, область применения.

Пример:

occupational safety and health management system	OSH management system	система управления охраной труда	СУОТ	набор взаимосвязанных или взаимодействующих между собой элементов, устанавливающих политику и цели по охране труда и достижение этих целей — A set of interrelated or interacting elements to establish OSH policy and objectives, and to achieve those objectives	ILO-OSH 2001 Guidelines on management systems for occupational health and safety // МОТ-СУОТ 2001 Руководство по системам управления охраной труда
ocean-bottom cable		донная (сейсмическая) коса			Морская сейсмозащита
ocean-bottom node	OBN	донный сейсмоприемник		Используются при сейсмозащите на море вместо донных сейсмических кос.	Морская сейсмозащита

<sup>1</sup> Распространяется как свободное программное обеспечение.

<sup>2</sup> Глоссарий – словарь узкоспециальных терминов. С программной точки зрения разницы между словарями и глоссариями нет.

Можно видеть, что все ячейки кроме первой ячейки в строке могут быть пустыми. Можно оставлять незаполненными целые колонки (кроме первой). Но в последней колонке должна быть заполнена хотя бы одна ячейка, чтобы число колонок с точки зрения программы было равно шести.

В одноязычном словаре должно быть ровно 5 колонок:

1. Термин.
2. Сокращенный или альтернативный вариант термина.
3. Определение, объяснение
4. Примеры употребления, дополнительные пояснения.
5. Источник, область применения.

Пример:

Daily Contract Quantity	DCQ	The amount of gas which a Buyer nominally undertakes to purchase and a Seller undertakes to deliver in a defined 24 hour period. Although featuring in many contracts, in practice this expression is of little meaning in itself. It may serve as a means of expressing the Annual Contract Quantity if the latter is expressed as a number of days multiplied by the DCQ. It may also define the rate at which the Seller must be able to supply gas.	See also: "Daily Delivery Rate".	Natural gas and LNG
Daily Delivery Rate		The Daily Delivery Rate (DDR) is the rate at which the Seller's facilities must be capable of delivering gas, expressed as a volume of gas per day, or as a multiple of the Daily Contract Quantity.	Also known as the Maximum Daily Quantity.	Natural gas and LNG
direct acting pressure relief valve		A pressure-relief valve that allows line pressure to act on one side of the closure element, while a predetermined spring applies a mechanical load to the other side. When the line pressure reaches its maximum limit, the line pressure overcomes the spring load and the valve opens until the line pressure falls below the preset level.	According to working principle, the safety valve or relief valve can be divided into direct-acting safety valve or relief valve and non-direct-acting safety valve or relief valve. Non-direct-acting safety valve or relief valve is divided into pilot operated safety valve or relief valve and safety valve or relief valve with auxiliary power device.	

Правила по заполнению ячеек те же, что и для двуязычных глоссариев.

В файле **не должно быть символов переноса строки и конца абзаца, т.к. это приведет к ошибкам при преобразовании типа!**

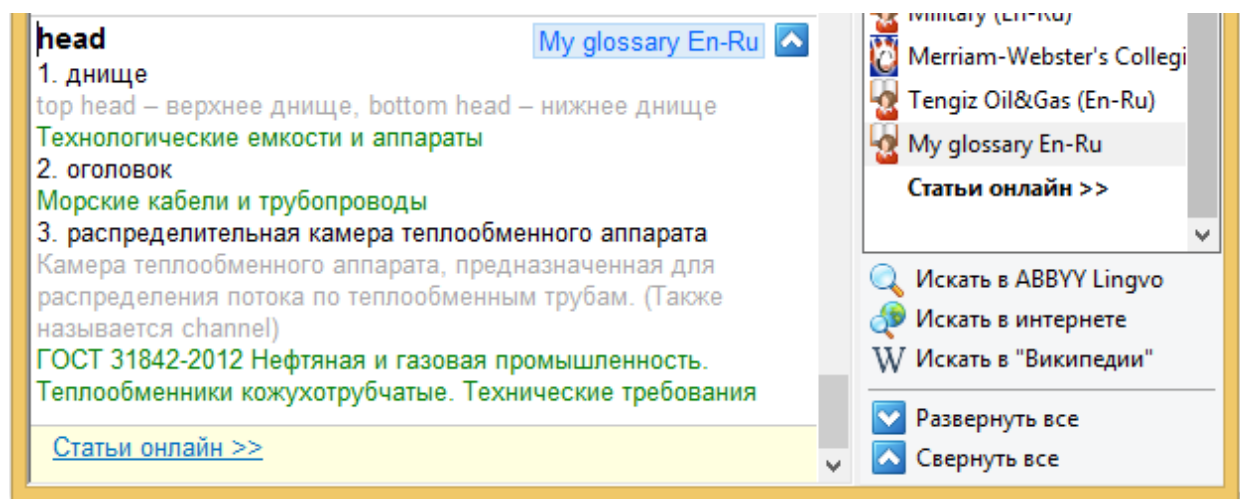
Кроме того, в файле **не должно быть квадратных скобок [ ]**, так как они используются при разметке словарей Лингво.

Если у одного термина должно быть несколько переводов или определений, необходимо создать несколько строк с одинаковыми первыми ячейками.

Пример:

head	1. днище	top head - верхнее днище, bottom head - нижнее днище	Технологические емкости и аппараты
head	2. оголовок		Морские кабели и трубопроводы
head	3. распределительная камера теплообменного аппарата	Камера теплообменного аппарата, предназначенная для распределения потока по теплообменным трубам. (Также называется channel)	ГОСТ 31842-2012 Нефтяная и газовая промышленность. Теплообменники кожухотрубчатые. Технические требования

Тогда в итоге будет получена словарная статья с тремя вариантами перевода:



Номера для разных переводов нужно проставлять вручную. В принципе можно этого не делать и просто добавлять новые строки в конец таблицы, не беспокоясь о том, что где-то уже есть строка для этого термина. Программа автоматически обработает такие строки и выдаст словарную статью с несколькими вариантами перевода, только они не будут пронумерованы.

Далее каждый лист глоссария надо сохранить в формате «Текстовый файл с разделителями табуляции».

### **Преобразование**

1. Запускаем программу на языке Python. Для этого нужно установить оболочку (Python IDE), которую можно скачать с сайта python.org. Запустив Python IDE, открываем файл с программой и вводим данные по ее запросу. В итоге получается файл с расширением DSL).
2. Файл нужно открыть в Wordpad и сохранить в формате Unicode (проследить, чтобы расширение файла осталось dsl, не txt).
3. Преобразовать в формат LSD (словарь Лингво) с помощью компилятора Лингво (DSL Compiler), который входит в комплект программ, поставляемых при покупке словарей Лингво.

## **Обоснование**

### **Выбор программ**

Рабочую версию своего глоссария веду в Excel, т.к. в этом формате его легко пополнять и редактировать. Периодически конвертирую в формат Лингво и подключаю как пользовательский словарь.

Почему не в базе терминов CAT? Да, в программах типа Trados и Mem-source кроме памяти переводов есть базы терминов. Но я не все переводы делаю в таких программах. Кроме того, я перевожу устно. Во время устного перевода не часто, но бывает необходимо заглянуть в словарь в самых разных условиях – и на буровой площадке и в кабине для синхронного перевода.

Лингво удобен быстрым доступом, а также возможностью полнотекстового поиска (не только по заголовочному слову, но и по другим частям словарной статьи). Недостаток: в версии для мобильных устройств подключение пользовательских словарей не предусмотрено, поэтому на смартфоне использую словарь GoldenDict, к которому можно подключать пользовательские словари в формате Лингво (lsd).

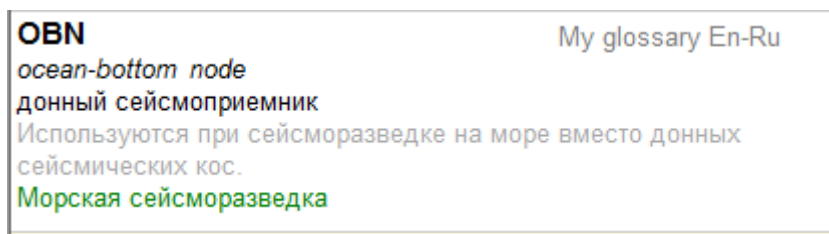
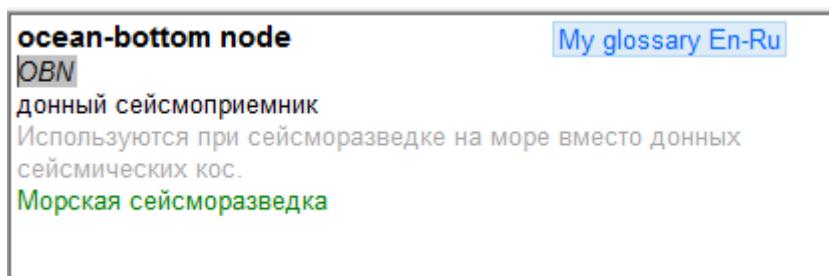
Python был выбран потому, что я сейчас изучаю этот язык ☺

### Состав словарной статьи

Термины могут иметь разный смысл и разный перевод в зависимости от области применения, заказчика и контекста. Для качественного перевода нужно понимать смысл употребляемых терминов, поэтому кроме термина и перевода в глоссарий включаю определение, источник, область применения и т.п.

Я также веду «одноязычные» глоссарии, где нет перевода, есть только толкования. Опыт показал, что в стандартах типа ГОСТ, ИСО есть много полезных терминов с определениями, что очень помогает в работе.

Если у термина есть сокращенный или альтернативным вариант (см. выше ocean bottom node – OBN), в итоговом пользовательском словаре в формате Лингво будет две статьи: в одной заголовочным словом будет термин, в другой заголовочным словом будет сокращение, а полный вариант будет альтернативной формой (см. ниже). Это упрощает поиск, особенно на мобильных устройствах, где поиск возможен только по заголовочному слову.



Казаков Денис Валерьевич. Переводчик

<http://denis-kazakov.com>