# One World, One Wiki!

C. Scott Ananian <cananian@wikimedia.org> [[User:cscott]] (Wikimedia Foundation)

## Language Converter

MediaWiki uses [[mw:LanguageConverter]] to automatically transliterate articles between closely related languages or dialects or script variants of a language or dialect. It is used on 11 wikis, and has been requested on about 35 more. Here are some examples of conversion pairs:

| | | |
|---|---|---|
| **English (American/British)**<br>*LanguageConverter not used.*<br>Spelling and usage differences exist between American English, British English, Indian English, and others. | [[**Elevator**]]<br>An **elevator** is a type of vertical transportation that moves people or goods between floors (levels, decks) of a building, vessel, or other structure. | [[**Lift**]]<br>A **lift** is a type of vertical transportation that moves people or goods between floors (levels, decks) of a building, vessel, or other structure. |
| **Serbian (Latin/Cyrillic)**<br>*LanguageConverter in use.*<br>Speakers are fully functionally digraphic, using both Cyrillic and Latin scripts. There are also vocabulary differences between Ikavian, Ekavian, and Ijekavian dialects which are not currently converted. | [[**Lift**]]<br>**Lift** је уређај за transport ljudi ili tereta među spratovima zgrada ili radnih platformi. Uobičajno se kreće uz pomoć elektromotora koji ili pokreće užad za vuču i protivtežni mehanizam, ili pumpa hidrauličnu tečnost za podizanje cilindričnih klipova. | [[**Лифт**]]<br>**Лифт** је уређај за транспорт људи или терета међу спратовима зграда или радних платформи. Уобичајно се креће уз помоћ електромотора који или покреће ужад за вучу и противтежни механизам, или пумпа хидрауличну течност за подизање цилиндричних клипова. |
| **Chinese (Simplified/Traditional)**<br>*LanguageConverter in use.*<br>Simplified used in mainland China, Singapore, and Malaysia. Traditional used in Taiwan, Hong Kong, Macau, and among Overseas Chinese. Most speakers monographic; few can fluently proofread text in both variants. | [[**电梯**]]<br>**电梯**，亦称升降机、垂直电梯。在香港、新加坡和马来西亚俗称“铊”（英语lift的译音），是一种垂直运送行人或货物的运输工具。 | [[**電梯**]]<br>**電梯**，亦稱升降機、垂直電梯。在香港、新加坡和馬來西亞俗稱「鉈」（英語lift的譯音），是一種垂直運送行人或貨物的運輸工具。 |
| **Hindi/Urdu**<br>*LanguageConverter not used.*<br>Urdu and Hindi are dialects of the Hindustani language, written in very different scripts: Arabic on the Pakistan side of the border, Devanagri on the India side. (Punjabi is a similar case, with four scripts used.)<br>Currently separate small wikis; could combine efforts. | [[**उत्थापक**]]<br>**उत्थापक**, उच्चालित्र अथवा एलिवेटर (lift या elevator) एक युक्ति है वस्तुओं एवं व्यक्तिओं को उर्ध्व दिशा में चढ़ाने-उतारने के काम आती है। प्रायः किसी बहुमंजिला ऊँचे भवन, जलपोत एवं अन्य संरचनाओं में उत्थापक लगा होता है जो गोलों को या सामान आदि को एक मंजिल से दूसरी मंजिल या एक स्तर से दूसरे स्तर पर लाता और ले जाता है। उत्थापक प्रायः विद्युत मोटर द्वारा चलते हैं। | [[**رافع**]]<br>انتصابی نقل و حمل کی کل ۔ جدید عمارتوں، ج۔ ازوں اور کانوں میں استعمال ۔ ون۔ والی تمام کھلی اور بند ساختوں اور لگاتار چلن۔ وال۔ ان پٹوں کو بھی رافع یا (انگریزی:Elevator) ک۔ ا جاتا ہے۔ جو بھاری چیزوں کو ایک جگ۔ س۔ دوسری جگ۔ پ۔ نچات۔ ۔یں۔ |

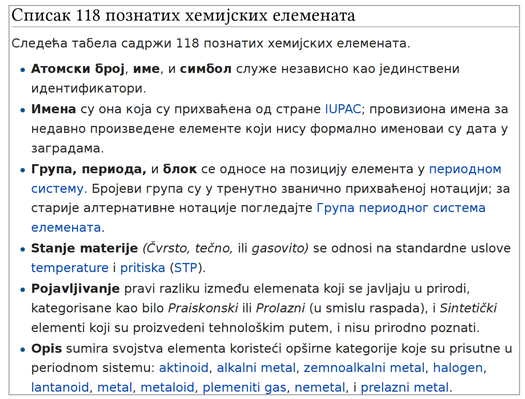
## Native Variant Editing

LanguageConverter is oriented to readers: it converts the article text unidirectionally into readable text in a consistent variant. But as soon as a user begins to edit, they are confronted with the source text in a mix of variants, as illustrated by the intermingled Cyrillic and Latin scripts in the article from Serbian Wikipedia shown below. This mixture of scripts can be a huge barrier to editing in communities where individuals are typically only fluent in a single variant.

The Parsoid team has been experimenting with a new bidirectional implementation of LanguageConverter, based on Finite State Transducers (FSTs). These allow automatic annotation of wikitext such that it can be round-tripped to its original variant losslessly. With these annotations, an Wikimedian can edit an article in their preferred consistent variant.

Unedited portions of the article will round-trip to their original variant, preventing dirty diffs, and only edited sections will reflect the variant which the editor saw. On wikis where the community has chosen to author all articles in a single variant, all text can be losslessly saved as the chosen variant, regardless of which variant the editor used.

**We can make editing easier on wikis using LanguageConverter!**

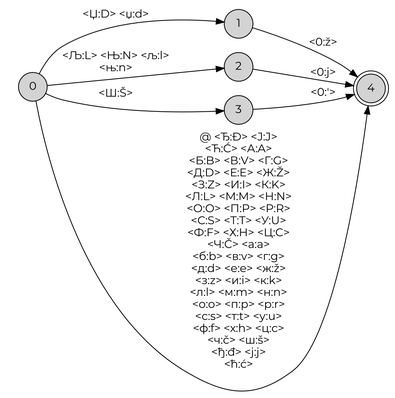


Списак 118 познатих хемијских елемената

Следећа табела садржи 118 познатих хемијских елемената.

- **Атомски број**, **име**, и **симбол** служе независно као јединствени идентификатори.
- **Имена** су она која су прихваћена од стране IUPAC; провизиона имена за недавно произведене елементе који нису формално именовани су дата у заградама.
- **Група, периода,** и **блок** се односе на позицију елемента у периодном систему. Бројеви група су у тренутно званично прихваћеној нотацији; за старије алтернативне нотације погледајте Група периодног система елемената.
- **Стање материје** *(Čvrsto, tečno,* или *gasovito)* се односи на standardne uslove temperature i pritiska (STP).
- **Појављивање** прави разлику између елемената који се јављају у природи, категорисане као било *Praiskonski* или *Prolazni* (у смислу распада), и *Sintetički* елементи који су произведени технолошким путем, и нису природно познати.
- **Opis** sumira svojstva elementa koristeći opširne kategorije koje su prisutne u periodnom sistemu: aktinoid, alkalni metal, zemnoalkalni metal, halogen, lantanoid, metal, metaloid, plemeniti gas, nemetal, i prelazni metal.

Serbian Wikipedia article showing mixed Cyrillic/Latin script




Partial FST for Serbian Cyrillic-to-Latin conversion

## Translation Suggestion Tool

lala

## Zero-Shot Translation

lala