

WMTrans / Canoo.net Case Study

LEO English-German Online Dictionary

Tuning Dictionary Lookup - How the LEO online dictionary profits from WMTrans Products

The Customer

With over 1.5 million queries per day LEO is the most frequently used English-German online dictionary. It can be accessed free of charge, worldwide. It is continuously maintained and extended by a team of computer scientists and linguists at the Technical University of Munich. Both in size and processing speed, the LEO dictionary is unparalleled in the Internet community. Due to the rapid growth in access rates, from 6 million to 38 million queries per month in the last two years, a dedicated dictionary server and a specialized maintenance concept has been set in place.

Functionally, the LEO dictionary has been enhanced and extended continuously. Today, it has reached a high level of sophistication:

- Corrections are suggested for possibly misspelled queries,
- Vocabulary-related issues can be discussed with other LEO users,
- Browser integration is available to increase ease of use,
- German inflected forms are recognized and i. reduced to base forms, and ii. linked to in-depth grammatical information.

The Challenge

Many users enter inflected word forms unaware that inflected forms are not listed. They enter the German word *bringst*, not knowing that this is an inflected form of *bringen* (to bring). Similarly, users often type outdated spelling variants like *numeriert* instead of *nummeriert* (numbered), which has been revised in the German spelling reform. Due to this user behaviour, there was a need to recognize inflected word forms and different spelling variants and to transform these queries into valid search terms, which then yield a result during dictionary lookup. In order to implement this feature, a large-scale morphological component was required, which had to fulfil the highest standards regarding processing speed, quality and volume.

Unfortunately, automated transformations of user input have negative as well as positive effects on usability, because they are not intelligible to the user. The English speaker entering *bringst*, for example, may not know that this is a form of *bringen*. For this reason, LEO needed explanatory information whenever the system carried out automated transformations. Background information enables the user to understand why the input was transformed, and thereby offers self-learning opportunities.

The WMTrans / Canoo.net Solution

WMTrans consists of a range of morphology software products to process text input. The components analyze individual words and, depending on the product:

- i. Determine the base form and both inflectional and spelling variant features,
- ii. Compute all possible inflected forms and spelling variants for a given base form,
- iii. Determine the components of a word from which it has been derived or composed, of
- iv. Determine all possible word compounds and derivations in which a given word is involved.

In the LEO scenario, the first of these functionalities was required to resolve the issue of recognizing and transforming German inflected forms and spelling variants. As a consequence, the WMTrans Lemmatizer product was chosen. This product is a very fast and lean finite state engine, providing the exact functionality required.

Canoo.net is an online service that offers a wealth of information on German grammar. Being based on the WMTrans suite of products and in-depth descriptions of both German inflection and the spelling reform, it delivers the type of explanatory information required for the LEO challenge described above. Since all query service and information pages at Canoo.net can be invoked via URL, a seamless integration into LEO was easy.

The Results

The enhancement of LEO's functionality using WMTrans and Canoo.net was one of the reasons why LEO was chosen Yahoo Site of the Year in December 2001.

LEO benefits from WMTrans and Canoo.net products in the following way:

Recognition of inflected query input

- The user types a query in LEO's search field
- If the number of entries in the dictionary is below a predefined threshold, the system checks whether the query is an inflected form and whether a spelling correction can be proposed. The first step is carried out by WMTrans.
- If WMTrans returns a base form that differs from the query, this form is searched for in the LEO dictionary.
- If the term is found in LEO, both the base form and possible spelling corrections are offered to the user for further selection.

Hyperlinking of query input and inflection tables

- In case of inflected user input, hyperlinks are generated, which refer to an inflection table showing how the query string can be derived from the base form. The inflection tables are available in the Inflection Dictionary of Canoo.net.

Hyperlinking of query result with inflection and orthographic information

- In case the query input or the query result is affected by the German spelling reform, a link to the respective entry in Canoo.net's Spelling Dictionary is displayed.
- If the query input corresponds to a base form in Canoo.net's Inflection Dictionary, links from the query's occurrences on the result page are generated, pointing to the respective entry in Canoo.net's Inflection Dictionary.

Hyperlinking of Canoo.net dictionary entries and LEO translations

- The Canoo.net applications profit from a back-link to LEO: if an English translation is available for a German word that has been entered at Canoo.net, a translation link is provided.















11 search results for 'zusammenschliessen'	
ENGLISH	GERMAN
to affiliate	zusammenschließen  
to combine	zusammenschließen  
to incorporate	zusammenschließen  
to integrate	zusammenschließen  
to merge [econ.]	zusammenschließen - von Unternehmen etc.  
to align with so.	sich mit jmdm. zusammenschließen  
to join forces	sich zusammenschließen
to confederate	sich zu einem Bündnis zusammenschließen
to syndicate	sich zu einer Interessengemeinschaft zusammenschließen
to cartelise chiefly [Brit.]	sich zu einem Kartell zusammenschließen
to cartelize chiefly [Amer.]	sich zu einem Kartell zusammenschließen

Figure 1: Sample Search Result for *zusammenschliessen*

Figure 1 shows a sample LEO search result. The button  directs the user to the word's inflection table and the button  links to entries in the spelling dictionary.

About Canoo

Canoo was founded in 1999 by a group of software engineers and senior managers that had previously been working together in the Web Technology centre of a major Swiss bank. Growing steadily while remaining committed to highest professional standards, Canoo has established itself as one of the top providers for complex Web applications in Switzerland. Canoo offers both IT services and a range of highly innovative products, including the linguistic products WMTrans and Canoo.net.

For more information on Canoo's WMTrans products, please consult:

<http://www.canoo.com/WMTrans>

or contact Canoo directly at:

Mail: wmtrans-info@canoo.com

Phone: +41 (61) 228 94 44

canoo.com

Canoo Engineering AG

Kirschgartenstrasse 7

CH-4051 Basel

Switzerland

Tel +41 (61) 228 94 44

Fax +41 (61) 228 94 49

<http://www.canoo.com/WMTrans>

wmtrans-info@canoo.com