

## WMTrans Products

### Multilingual Morphology Software - Word Analysis and Generation

WMTrans (WordManager Transducer) consists of a range of morphology software products to process text input. The smart text processing software is used in retrieval and language processing applications. Typical use cases include word stemming, intelligent search, text indexing, text mining, language learning, hyperlink generation, spell checking, grammar checking, and machine translation.

WMTrans is based on the Canoo Morphological Dictionaries, containing:

- more than 250'000 lexemes, generating 3 million fully categorized word forms for German,
- 50'000 lexemes, generating 115'000 word forms for English,
- and 50'000 lexemes, generating 456'000 word forms for Italian.

The dictionaries include information on word formation dependencies, all types of morphological irregularities and spelling variants.

In addition, WMTrans processes English and Italian contractions such as *can't*, *he'll*, *mother's*, *dimmi*, *credermelo*, and *pensandoci*, increasing the number of recognized word forms to over 200'000 for English and over 2 million for Italian.

#### Product Overview

WMTrans allows developers to integrate functions such as word stemming, spell checking, and paradigm generation into processing systems. The following products are available:

##### Lemmatizer

The Lemmatizer retrieves the base form and category of any valid word.

##### Inflection Analyzer

The Inflection Analyzer determines the base form and category of a word, plus further inflectional information.

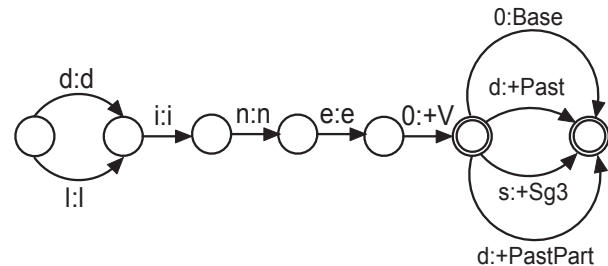


Figure 1: Example of a Finite State Transducer for words such as *dine*, *line* and variants

##### Recognizer

The Recognizer detects if a character string is a valid word and reports the result.

##### Generator

The Generator returns all inflected word forms and spelling variants for a base form.

##### Inflection Analyzer/Generator

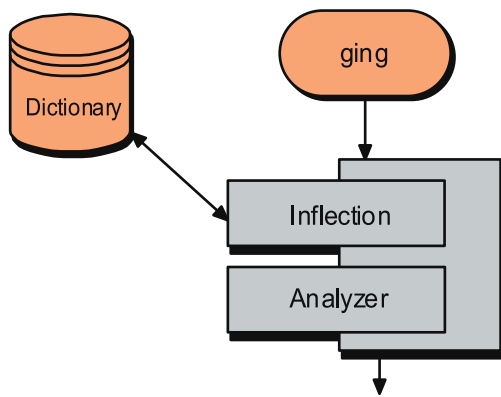
The Inflection Analyzer/Generator determines the base form and category of a word and computes all possible inflected forms and spelling variants for a given base form.

##### Word Formation Analyzer/Generator

The Word Formation Analyzer/Generator determines the components of a word from which it has been derived or composed and finds all possible word composites and derivations in which a given word is involved.

##### Unknown Word Lemmatizer

The Unknown Word Lemmatizer retrieves the base form, the category and the underlying word formation rule of a German compound or derived word not listed in the lexicon. Also available: Unknown Word Recognizer and Analyzer.



gehen  
 (Cat V)( Aux sein )(Mod Ind)(Temp Impf)  
 (Pers 1st)(Num SG)(ID 0-1),  
 (Cat V)( Aux sein )(Mod Ind)(Temp Impf)  
 (Pers 3rd)( Num SG)(ID 0-1)  
 #

Figure 2: Simplified Diagram showing the WMTrans Inflection Analyzer for German

## Benefits

### *High Optimization*

WMTrans uses finite-state transducers, a proven technology for word processing. Benefits include runtime efficiency and low memory requirements.

### *Modular Setup*

All WMTrans products are modular. No need to buy the full software product range, if a single program solves the issue.

### *Easy Integration*

Using the API for Java or C/C++, the required functionality can be easily added to a Java or C/C++ application. The WMTrans products can be integrated as stand-alone programs (data compilers) or shared libraries (analyzers/generators) into existing or new applications.

Customer data can be integrated using WMTrans compilers. On demand, Canoo offers consulting and support at any project stage.

Applying WMTrans will save development time and money. Time to market is reduced. Developers can concentrate on building the application and rely on WMTrans products to compute the relevant data.

## Test WMTrans Now

Download the free evaluation licence at:

<http://www.canoo.com/wmtrans/download>

Browse through the product description, test the APIs and find out how functions can be integrated into your application.

## See WMTrans in Action

Canoo Online Services are based on WMTrans Products and provide examples of possible applications.

The online services can be tested at:

<http://www.canoo.net>

## Supported Languages

- German
- English
- Italian

## Supported Operating Systems and Compilers

- Linux with Java
- Linux with ANSI C++
- Windows NT (in preparation)

## Supported Spelling Systems

- New and old German spelling
- German, Swiss, and Austrian variants
- NZZ , FAZ, dpa, Die Zeit, and Süddeutsche Zeitung variants

## 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.

**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](mailto:wmtrans-info@canoo.com)