

Integrating Linguistic Products into Corporate Applications

Elisabeth Maier

Canoo Engineering AG

Basel-Switzerland

elisabeth.maier@canoo.com

www.canoo.com, www.canoo.net

Outline

- Background
- Canoo's Building Blocks
 - Content
 - Technologies
 - Services
- Sample Applications

Background

- Canoo operational since January 2'000
- Spin-Off
 - Basel University
 - UBS AG
- Licensing of research and products
 - Word Manager morphological database and technology (Basel University)
 - ULC – Rich Clients für J2EE (IBM)
- High-end company profile
 - excellent delivery track record
 - significant industrial experience
 - over 30% of staff with PhD
 - Ability to execute “in the large”

Canoo's Content (I)

Word Manager Morphological Databases of

- German (currently > 250'000 base entries)
- Italian (> 50'000 base entries)
- English (> 50'000 base entries)

Information

- word category: i.e. noun, adjective, verb, etc.,
- word form features: number, case, person etc.
- word formation information (word formation rules, base words / units, segmentation)
- spelling variants, including
 - spelling rules variants introduced due to spelling reform of German
 - regional variants (Swiss German, Austrian German; American vs. British English)
 - publisher-specific variants (e.g. Neue Zürcher Zeitung, Frankfurter Allgemeine Zeitung, die Süddeutsche, die Zeit, Deutsche Presseagentur)

Canoo's Content (II)

Canoo Online Dictionaries

- Spelling Dictionary
- Inflection Dictionary
- Word Formation Dictionary
- Glossary of Linguistic Terms

Static Grammar Pages – Textbook Knowledge for Learners of German

- Over 1'700 platform-independent HTML pages
- Interactive links to Canoo Dictionaries
- Information on
 - Spelling, Inflection, Word Formation, Grammar
- Incorporating Metasearch

Canoo's Technologies (I)

WM Transducer Framework

- Object-oriented framework for the automatic generation of transducers according to input-/output specifications
- Creation of
 - Standard products
 - Customer specific tools

WMTrans Products

- Transducer-based technology for processing WordManager data
- Extensive product range including
 - Recognizers, lemmatizers, analyzers and generators for lexicalized words
 - Unknown Word products (recognizers, lemmatizers and analyzers) for non-lexicalized words
- Multi-platform implementation (pure Java); partially also shared library implementation available (Linux OS; C++ and Java API)
- Available for German, Italian and English

Canoo's Technologies (II)

- ULC – Rich Clients for J2EE
 - Rich GUI library for fast and cost efficient development of lightweight clients
 - Solution for centralized server-side programming and application deployment - generating minimal network traffic
 - Ideal for
 - Ambitious thin client GUIs where HTML is not sufficient
 - Handling & display of large amounts of data
 - Realization of Thin Client Multichannel Architecture (ULC & HTML Integration)
- WebApp
 - Object-oriented application framework for the development of scalable Servlet- and JSP-based Web applications.
- WebTest
 - free open source tool for automated testing of web applications

Canoo's Services

- Software development
- Software integration
- Technical consulting, reviews
- Project management (development projects)
- On-the-job Training
- ASP for applications developed / integrated by Canoo

Sample Applications

E-Learning

Canoo.Net, Online Dictionaries and Grammar for German

- Standard GUI
- Incorporating ULC Search
- „HTML“-Integration of Content into 3rd party applications
- Free Online Service with 10'000 unique visitors/day



Powered by Canoo Products

[Site Map](#) | [Help](#) | [Subscribe](#) | [Contact](#) | [Feedback](#) | [Deutsch](#)

 Search

[Dictionaries](#) | [Word Grammar](#) | [Word Formation \(D\)](#) | [German Orthography \(D\)](#) | [Glossary](#) | [Index](#)

German Dictionaries and Grammars

[Interactive Examples](#)

[High-End-Search with ULC User Interface](#)

Mann

search

in: All

 case sensitive ss instead of ß[Deutsch](#) | [Sitemap](#) | [Help](#)[Registration](#) | [Contact](#) | [Feedback](#) | [Links & Bookmarks](#)[Dictionaries](#) | [Word Grammar](#) | [Wordformation](#) | [German Orthography \(D\)](#) | [Glossary](#) | [Index](#)

Inflection of *Mann*

[Spelling](#) | [Inflection](#) | [Wordformation](#) | [Morphology Browser](#) | [LEO \(D-E\)](#)[Fully inflected](#) | [Stem+Ending](#)**Word class:** [Noun](#)**Genitive singular:** Mannes / Manns**Nominative plural:** Männer**Article:** der**Inflection class:** [\(e\)s/er](#)**Features:** [plural with umlaut](#), [genitive with s or es](#), [dative e](#)

	Singular		Plural 1	
	Article	Noun	Article	Noun
Nominative	der	Mann	die	Männer
Accusative	den	Mann	die	Männer
Dative	dem	Mann Manne	den	Männern
Genitive	des	Mannes Manns	der	Männer

Mann search
in: All case sensitive ss instead of ß

Dictionaries

- Mann: Noun, Masculine [Spelling](#) [Inflection Dictionary](#) [Wordformation Dictionary](#) [Morphology Browser](#) [LEO \(D-E\)](#)
- Mann: Noun, Masculine, Reading Gefolgsmann [Spelling](#) [Inflection Dictionary](#) [Wordformation Dictionary](#) [Morphology Browser](#) [LEO \(D-E\)](#)

Männergesangverein search
in: All case sensitive ss instead of ß

The Canoo dictionary does not contain your input word.

Using the [Unknown Word Analyzer](#) your input can be decomposed into the following known entries:



search

in: Grammar case sensitive ss instead of ß

- Verb (D)
- ▼ Inflection
 - Introduction
 - Inflection classes
 - ▼ Inflection tables
 - ▼ Simple tenses
 - Present
 - Modal verbs
 - Preterite
 - Regular
 - Irregular**
 - Modal verbs
 - Imperative
 - Modal verbs
 - Infinitive and participle
 - Past participle
 - Modal verbs
 - Complex tenses
 - Tense-mood signal + endings
 - Features
 - Grammatical categories (D)
 - Semantic classes (D)
 - Aspect (D)
 - Full, modal and auxiliary verbs (D)
 - Verb and other parts (D)

Irregular preterite

[Rule](#) | [Stem+Ending](#) | Fully inflected

Preterite			
Indicative		Subjunctive II	
Person	Verb	Person	Verb
ich	sang	ich	sänge
du	sangst	du	sängest
er/sie/es	sang	er/sie/es	sänge
wir	sangen	wir	sängen
ihr	sangt	ihr	sänget
sie	sangen	sie	sängen

Preterite indicative

The irregular forms of the preterite indicative are composed of the preterite stem (past stem) and the person endings of the preterite.

Features:

- ◆ [e-addition](#) in the 2nd person plural
- ◆ [s-merging](#)
- ◆ Mixed conjugation (see below)

Subjunctive II

In the formation of the irregular forms of the subjunctive II the preterite stem is unlauded whenever possible. It is combined with

Kreditgeschäft **Suchen**

Kreditgeramung
Kreditgenossenschaft
Kreditgeschäft
Kreditgewährung
Kreditthai
Kredithilfe
kreditieren
Kreditieren
kreditierend
kreditiert
Kreditierung
Kreditinstitut
Kreditiv
Kreditkarte
Kreditkauf
Kreditlinie
Kreditmarkt
Kreditnehmer
Kreditnehmerin
Kreditor
Kreditorenkonto
Kreditplafond
Kreditpolitik
kreditpolitisch
Kreditrestriktion
Kreditschutz
Kreditseite
Kreditsicherung
kreditunwürdig
Kreditunwürdigkeit
Kreditversicherung
Kreditvolumen
Kreditwesen
kreditwürdig

Powered by [Canoo Products](#)

search

[Deutsch](#) | [Sitemap](#) | [Help](#)

in: All

 case sensitive ss instead of ß[Registration](#) | [Contact](#) | [Feedback](#) | [Links & Bookmarks](#)[Dictionaries](#) | [Word Grammar](#) | [Wordformation](#) | [German Orthography \(D\)](#) | [Glossary](#) | [Index](#)

Dictionaries

Kreditgeschäft: Noun, Neuter [Spelling](#) [Inflection Dictionary](#) [Wordformation Dictionary](#) [Morphology Browser](#) [LEO \(D-E\)](#)Copyright © 2000-2003 [Canoo Technology AG](#), Kirschgartenstr. 7, CH-4051 Basel, tel. +41 61 228 94 44. All Rights Reserved.

Sample Applications (II)

Machine Translation

MT Infrastructure for Corporate Language Services AG

- Evaluation of MT products
- Setup and extension of MT system to serve diverse customers with different requirements
- Development and maintenance of infrastructure (e.g. authentication, logging, monitoring, and accounting)



CLS Corporate Language Services AG

CLS Machine Translation

Deutsch **English** Français Español

Text

Texts consisting of up to 4,096 characters can be inserted into the text field. A text can be hand-typed or transported from the clipboard by using the copy and paste function. Select the translation direction and subject area. Click on *Translate* to send your request to **CLS Machine Translation**. How long the translation takes depends on the level of system utilization.

Document

Translation direction:

German-English ▾

Support

Subject area:

Telecommunications ▾

FAQ

Show alternative translations:

No ▾

Help

Text:

Die Leitweglenkung zu einer verketteten Nummer unterliegt den gleichen Anforderungen wie zu der in der verketteten Nummer enthaltenen Nationalen Nummer.

Translate

The translated text can be viewed below.

Translation:

The alternative routing for a concatenated number is subjected to the same requirements as for the National number contained in the concatenated number.

Sample Applications (III)

IR & Text Mining

Google

Polderland

Catalogue & Dictionary Lookup

LEO's Online Dictionary for German

- Stemming functionality for query processing
- Link to inflection tables

ELDIT Italian-German Online dictionary

Other NLP applications

Spell checker for NZZ Online



LEO - Link Everything Online

An Online Service by [Informatik der Technischen Universität München](#)



Search term (English or German):

wasserfall

Search

Clear

[Usage Hints](#)

Computer DVD-Filme Österreich
 Personal HIFI Preis finden Deutschland



Advertisement

2611986 requests today

This dictionary contains 394791 entries!

7 search results for 'wasserfall'

Options: (Default bold)

Direction:
 English > German
 German <-> English
 German > English

Link Results:
 No Yes

Spelling Tolerance:
 Switched off Force
 Standard

Headlines:
 Off On

Presentation of Results:
 With Grid Without

Special Characters Tolerance:
 High Switched off
 Standard

Choose Language:
 German English

ENGLISH		GERMAN	
Direct Matches			
m	cascade	der Wasserfall	g d P
m	cataract [geog.]	der Wasserfall	g d P
m	waterfall	der Wasserfall	g d P
Verbs and Verb Phrases			
	to talk one's head off	reden wie ein Wasserfall	g P
	to run off at the mouth [coll.]	wie ein Wasserfall reden [coll.]	g
	to talk nineteen to the dozen	wie ein Wasserfall reden	g P
Examples			
	He talks his head off.	Er redet wie ein Wasserfall.	

Orthographically similar words - German:

- ♦ [wasserball](#)



Legend

- d** Definition, German 4)
- e** Additional Information on 'Phrasal Verbs', English 1)
- f** Forum: Discussion
- g** Grammar, German 3)
- m** Pronunciation and Definition, English 2)
- o** Orthography, German 3)
- p** Pronunciation, German 5)
- v** Variant, English and German

1): Based on Englishpage.com
 2): Based on Merriam-Webster's Collegiate(R) Dictionary
 3): Based on Canoo.com
 4): Based on DWDS, Wörterbuch der Gegenwartssprache
 5): Based on PROSER by ATIP-GmbH