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 f
 ür 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

02/12/2003

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 Servletand JSP-based Web applications.

WebTest

free open source tool for automated testing of web applications

Canoo's Services

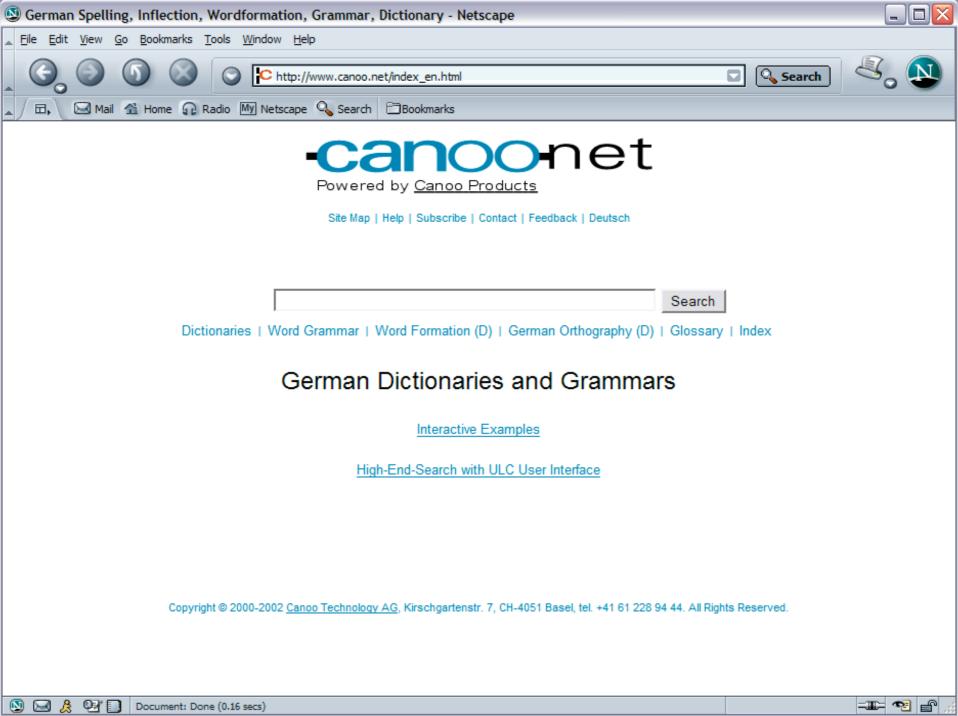
- Software development
- Software integration
- Technical consulting, reviews
- Project management (development projects)
- On-the-job Training
- ASP for applications developed / integrated by Canoo

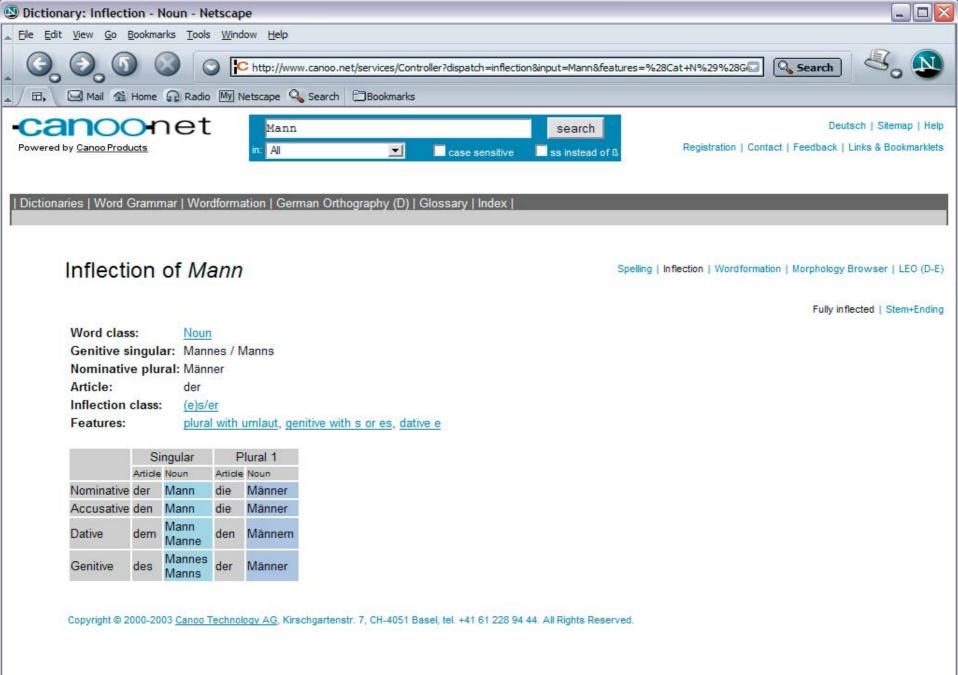
Sample Applications

E-Learning

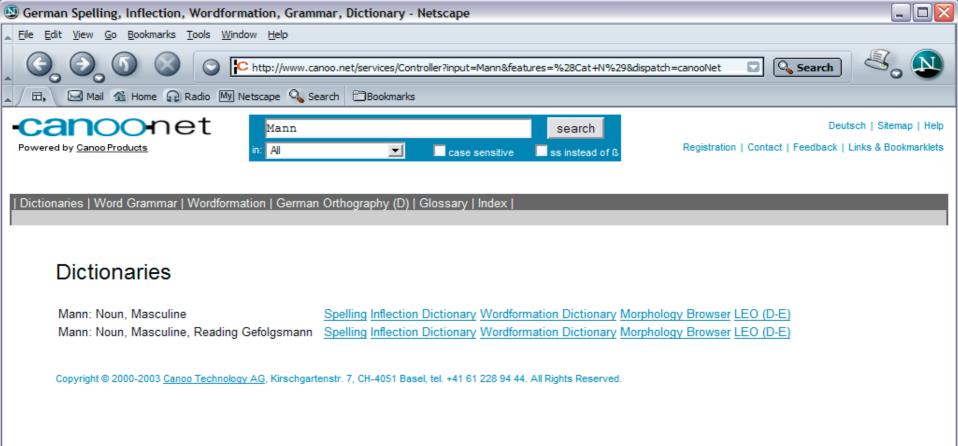
Canoo.Net, Online Dictionaries and Grammar for German

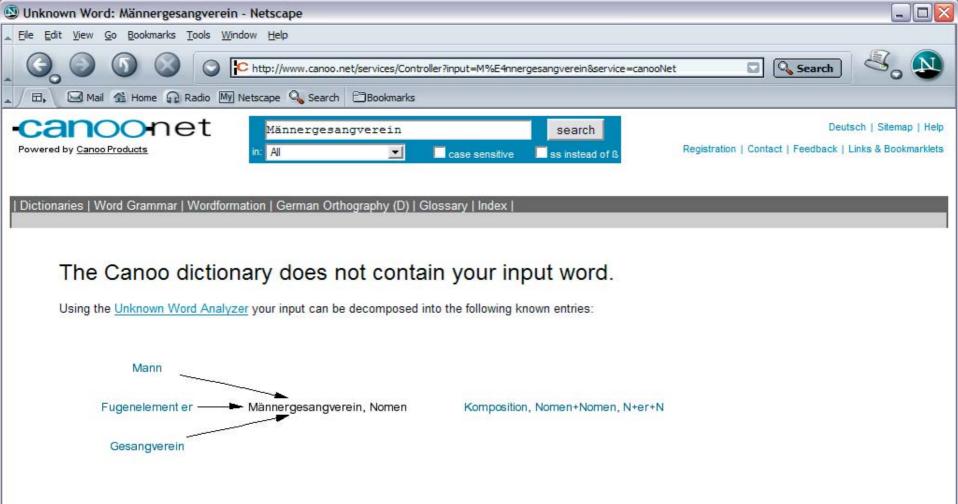
- Standard GUI
- Incorporating ULC Search
- "HTML"-Integration of Content into 3rd party applications
- Free Online Service with 10'000 unique visitors/day











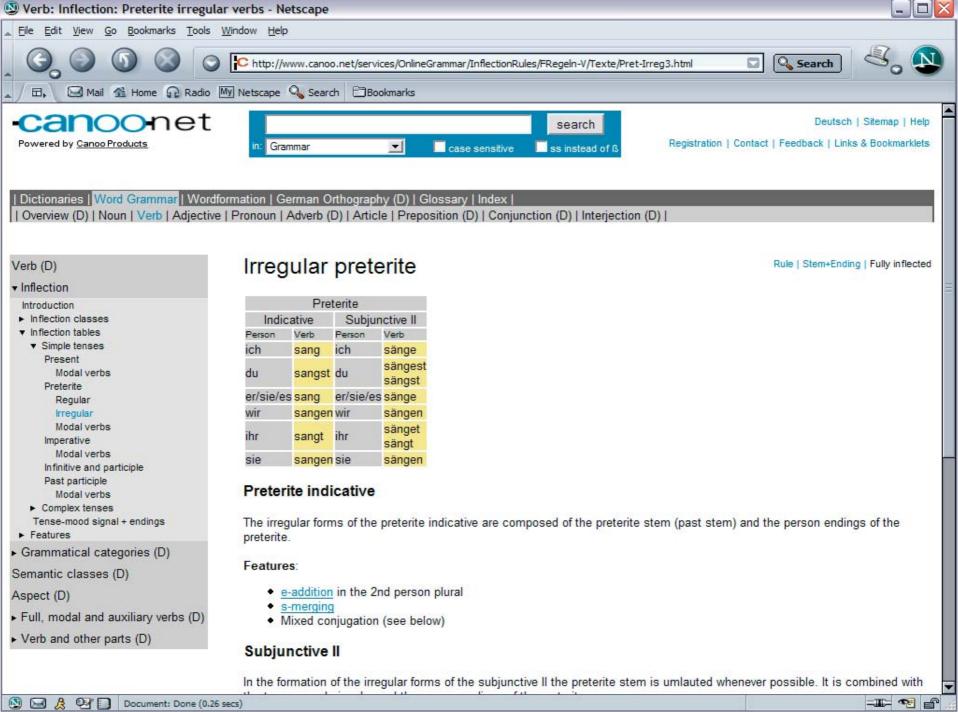
Copyright @ 2000-2003 Canoo Technology AG, Kirschgartenstr. 7, CH-4051 Basel, tel. +41 61 228 94 44. All Rights Reserved.

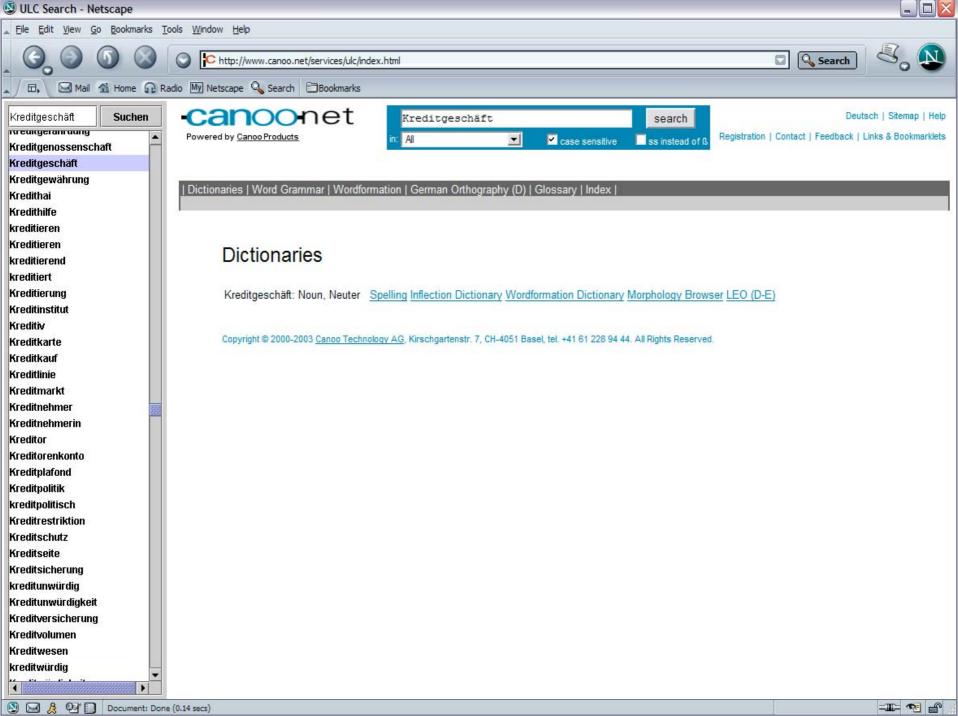












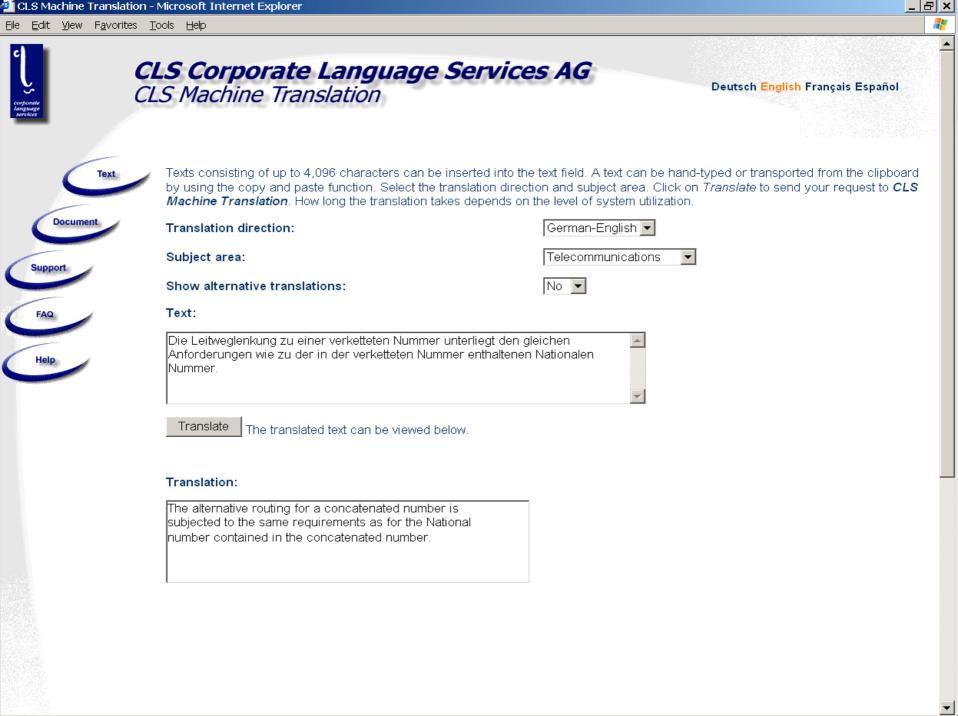
Sample Applications (II)

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)

02/12/2003



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

