



**VIRTUS™**

## **Virtus: Translation for Structured Data**

**Mark Arehart, [mark.arehart@ntrepidcorp.com](mailto:mark.arehart@ntrepidcorp.com)  
Ntrepid Corporation  
[www.virtustranslator.com](http://www.virtustranslator.com)**

### **Description**

This paper presents Virtus Translator<sup>1</sup>, a new machine translation (MT) product developed by Ntrepid Corporation. Virtus is a specialized tool designed for maximum performance on structured data, such as found in spreadsheets and database tables. The Virtus engine uses statistical machine learning techniques to identify the language and category of each column in a table. The categories for which Virtus is optimized include names of people, locations and organizations as well as supplemental information such as job title, education level, nationality and religion. Based on the automatically detected language and category, Virtus recommends a translation strategy suitable to maximize performance on each type of data.

Virtus translation strategies represent an innovation in translation software that provides users unprecedented control over the handling of their structured data. Virtus strategies range in sophistication from a "do not translate" approach suitable for internet addresses and certain numeric data, to a range of linguistically appropriate translation techniques. For example, a "transliteration only" strategy might be used for person names, while street addresses would be better handled by keyword translation supplemented by automatic transliteration. More sophisticated translation strategies are recommended for more linguistically complex categories such as organization names and job titles. Knowing the category of the data in each column allows not only for intelligent algorithm selection, but also enables more accurate translation of category-specific terminology, including acronym and abbreviation expansion.

The Virtus translation engine can run in fully automatic mode; however, it is also highly customizable, allowing users to specify translation parameters for each column of data, controlling the language, the category, the translation strategy and the transliteration standard on a per-column basis. Also, recognizing the uniqueness of each user's translation requirements, Virtus provides a custom translation memory (TM) building tool to allow sophisticated users fine-grained control over Virtus Translator's output.

Virtus Translator for Chinese, Persian and Russian are currently available. Virtus Translator for Arabic, French, German and Spanish are scheduled for release in 2012. A 30-day trial version is currently available upon request.

---

<sup>1</sup> Patent Pending 12/461,574