

# SYSTRAN Enterprise Server 7 Training Server

## Product Overview

SYSTRAN Training Server allows corporate users and language service providers to independently train SYSTRAN Enterprise Server 7 to any domain or business objective to produce high quality translations. As a result, SYSTRAN Enterprise Server 7 automatically learns from existing and validated translations and updates itself as these translations are reused.

SYSTRAN Training Server has two components. SYSTRAN Corpus Manager helps users acquire, manage and store parallel texts, existing translations and all other language assets. SYSTRAN Training Manager lets users run iterative training cycles based on predefined training tasks to fine-tune translation quality. The training tasks include resource extraction, dictionary validation and document alignment.

## Key Benefits

**Customize translations to a specific industry, corporation or subject.** Train the software to understand the terminology of a particular industry, business objective or any domain. Translation Models are created automatically to reduce customization time and cost.

**Leverage and reuse language assets.** A common repository enables the centralized storage and management of all existing language assets including user-defined dictionaries, Translation Memories, glossaries, previous translations, and all monolingual or multilingual data. All language assets can be reused to reduce costs of ongoing and new translation projects, dictionary development and maintenance.

**Work collaboratively.** Leveraging the common repository, multi-user efforts are supported. The advanced dictionary management tool and centralized repository of multilingual data allow users to share corporate language assets and manage work in a way that is systematic and easy to follow.



SYSTRAN Enterprise Server 7 - Copyright 2010 SYSTRAN all rights reserved

Id	Name	Status	Task Summary	Progress	GTM	WER	BLEU
6	IFRS_ENFR_Hybrid3_with_UD_Protected	Completed	Type Hybrid Training EN->FR Profile: IFRS Finance with UD Start: 2010-04-28 09:37:20 End: 2010-04-28 21:43:09	100	64.29	32.54	59.85
4	IFRS_ENFR_Hybrid2_with_UD	Completed	Type Hybrid Training EN->FR Profile: IFRS Finance with UD Start: 2010-04-27 22:55:38 End: 2010-04-28 08:34:52	100	63.95	33.22	60.12
3	IFRS_ENFR_Hybrid1	Publish Completed	Type Hybrid Training EN->FR Profile: IFRS Finance Default Start: 2010-04-27 22:53:56 End: 2010-04-28 07:55:42	100	64.36	30.71	60.43

## Key Features

**Manage parallel texts efficiently.** Store and manage all monolingual and multilingual documents, and Translation Memories. The language data is stored as translation units (TUs). Each translation unit consists of a source language segment and one or several target language segments. Individual files are uploaded into Corpus Manager and transformed into TUs. Corpus Manager maintains reference to the original files from which the TUs are extracted. Supported file formats are TXT, HTML, DOC, DOCX, XLSX, PPTX, PDF, TMX, and it is possible to support any XML format through customization.

**Train the translation engine.** Launch and supervise all automatic training tasks. Easily analyze and compare all training results.

Training tasks include:

- ▶ Generate baseline translations with automatic quality scores (BLEU, GTM, TER, Throughput, WER)
- ▶ Build Hybrid Translation Models to produce consistent quality translation results, leveraging both rule-based MT and statistical techniques
- ▶ Build Statistical Translation Models to produce quality translations using statistical techniques
- ▶ Create User Dictionaries with automatic terminology extraction from existing translations
- ▶ Validate User Dictionary entries with existing parallel texts
- ▶ Build a Translation Memory from existing parallel texts

## Training Server Configuration Requirements

### Operating System

Red Hat Enterprise Linux 5 (5.4 or later) and 6, 64-bit  
Linux CentOS 5 (5.4 or later) and 6, 64-bit

### Processor (Linux) \*

Minimum: Dual Quad-Core Intel Xeon or compatible (2 GHz or higher)  
Recommended: Dual Six-Core Intel Xeon or compatible (2 GHz or higher)

### Memory \*

Minimum: 16 GB RAM  
Recommended: 32 GB RAM or higher

### Hard Disk \*

Minimum: 120 GB free disk space  
Recommended: 250 GB+

### Server and RDBMS

Apache Tomcat 7.0 or later - SYSTRAN Training Server installs Tomcat 7.0 by default  
JDK 7 is required, and is included with SYSTRAN Training Server  
MySQL 5.5 is required

*\*Actual CPU/memory/hard disk requirements to be determined by the expected activity. Please speak with your Sales Account representative for more information.*

## Automatic Scores

### BLEU

#### BiLingual Evaluation Understudy

Compares MT output against a reference (human) translation. With a high quality training corpus, statistical machine translation would score above 50. A human translator is scored at around 80.

### GTM

#### General Text Matcher

Measures the similarities between machine translation output and a reference (human) translation. Some believe it correlates better with post-editing.

### TER

#### Translation Error Rate

Measures the number of edits required to change a system output into the reference translation.

### Throughput

#### Words/Second

Is calculated on the server hardware on which a model is trained and is used to compare different trainings run on the same server. This is useful to determine the optimal balance between performance and quality.

### WER

#### Word Error Rate

Provides the distance between the reference translation in a test corpus and a translation from a trained system. It is computed as the percentage of modified words in a translation unit between machine translation and the reference translation; the lower the score, the less post-editing is required.

NOTE: BLEU, GTM, TER, Throughput, WER, and other metrics do not strictly evaluate the quality of machine translation, but are useful in estimating the relative quality of comparable training tasks.

The screenshot displays the SYSTRAN Enterprise Server 7 web interface. The top navigation bar includes links for Translation, Dictionary Search, Dictionary Management, Translation Model, Corpus Manager, and Training Manager. The main content area is divided into three sections: Task Launcher, Current Activity, and Task Monitoring. The Task Launcher section on the left lists various tasks such as Baseline Evaluation, Hybrid Training, Statistical Training, Resource Extraction, Dictionary Validation, and Document Alignment. The Current Activity section shows a table of tasks with columns for ID, Name, Status, Task Summary, Progress, GTM, WER, and BLEU. The Task Monitoring section at the bottom provides a detailed view of the selected task, including its progress and quality scores.

ID	Name	Status	Task Summary	Progress	GTM	WER	BLEU
6	IFRS_ENFR_Hybrid3_with_UD_Protected	Completed	Type: Hybrid Training EN->FR Profile: IFRS Finance with UD Start: 2010-04-28 09:37:20 End: 2010-04-28 21:43:09	100	64.29	32.54	59.85
4	IFRS_ENFR_Hybrid2_with_UD	Completed	Type: Hybrid Training EN->FR Profile: IFRS Finance with UD Start: 2010-04-27 22:55:38 End: 2010-04-28 08:34:52	100	63.95	33.22	60.12
3	IFRS_ENFR_Hybrid1	Publish Completed	Type: Hybrid Training EN->FR Profile: IFRS Finance Default Start: 2010-04-27 22:53:56 End: 2010-04-28 07:55:42	100	64.36	30.71	60.43



SYSTRAN SA  
5 rue Feydeau  
75002 Paris  
FRANCE

SYSTRAN Software, Inc.  
4445 Eastgate Mall, Suite 310  
San Diego, CA 92121  
USA

For more information visit  
[www.systransoft.com](http://www.systransoft.com)