TRANSPERFECT

GlobalLink TM Server

Server-Side Translation Memory Technology

White Paper

November 2006 | Copyright TransPerfect 2006

1 Executive Summary

1.1 Introduction

The GlobalLink TM Server is the latest addition to the GlobalLink technology platform, and is scheduled to for general release in early 2007. GlobalLink TM Server applies a next-generation approach to translation memory (TM) management and virtual linguistic team collaboration, providing substantial advantages over traditional translation memory technology. This paper outlines the system's architecture at a conceptual level and provides an overview of key features and benefits, system requirements, and strategies for deployment.

GlobalLink TM Server is an enterprise application designed to provide maximum possible content re-use and facilitate collaboration between globally distributed localization resources. Architected for maximum extensibility, the GlobalLink TM Server can be deployed as a stand-alone application or integrated with the GlobalLink Content Director, GlobalLink Portal, and non-GlobalLink technologies.

1.2 Why Server-Based Translation Memory?

Despite the rapid uptake of TM technology in the mid to late 1990s, performance and scalability improvements in traditional TM applications failed to meet ever more demanding market requirements and were largely unable to take advantage of the Internet as a means of workgroup collaboration.

Server-based TM technology evolved to meet the needs of customers with substantial investments in TM assets and who had outgrown the capacity of these applications. Without a server-based translation memory solution, users with large-scale translation initiatives would either:

- Engage in the manual exchange of translation memories through email or FTP and lose the re-use benefit achieved by leveraging concurrent translator work, since translation memories would need to be gathered centrally, updated, and redistributed.

  Or:
- Depend upon a relationship with an outside language service provider for the management of translation memory assets and collaboration benefits, which can result in unacceptably high switching costs and unwanted barriers to exit from the service relationship.

A centralized approach to collaboration in a large-scale translation environment avoids both of these pitfalls and fosters substantially increased translation productivity and decreased time-to-market for localized content.

2 User Requirements & System Capabilities

2.1 Share Translation Memory Assets

GlobalLink TM Server users are seeking an environment in which groups of translators can drive productivity by sharing translation memory assets over the Internet or within a local area network environment.

2.1.1 Real-Time Collaboration on an Active Project Level

For maximum productivity, sharing must occur in real-time on a project level. This means that multiple translation resources concurrently assigned to a given project must easily be able to leverage one another's work.

2.1.2 Exercise Managerial Control Over TM Assets at an Enterprise Level

While GlobalLink TM Server users are interested in the maximum possible leverage, they are also aware that not all translation resources in a distributed work environment will be created equal. Technology must not only allow real-time sharing of translation memory assets by translators at a project level, but must also allow management to exercise controls over the inputs to translation memory at an enterprise level. Prior to 'accepting' a given translation memory unit produced in real-time, users will typically require some level of approval process so that their enterprise level TM remains up-to-date and properly managed at all times. Non-approved matches in the memory should still be made available for leverage, with a matching penalty deducted to ensure that the translator is aware of its status and performs the necessary proofreading.

2.1.3 Collaborate among a Large Pool of Translator Resources

GlobalLink TM Server users require a system that allows many translator resources to access translation memory assets simultaneously. There can be no measurable deterioration of performance as the number of translator resources requiring simultaneous access grows larger.

2.2 Work with Large-Scale Translation Memories

GlobalLink TM Server users desire an environment that allows for translation memory functionality to operate effectively and rapidly, even when leveraging from an enterprise level TM where individual assets may number in the millions. Users require Google-like performance in applying massive amounts of underlying data to traditional translation memory functionality.

2.3 Perform Localization Related Managerial Functions

GlobalLink TM Server users are required to perform a variety of functions that seek to maximize not only translation leverage and re-use, but also automation with respect to the entire content life cycle.

2.3.1 Translation Memory Management

GlobalLink TM Server users typically need to manage an enterprise level or centralized translation memory. These activities may include managing user rights and privileges with respect to updating the enterprise level TM, managing the locations and connectivity options for both the enterprise level TM and any departmental TMs, and general TM maintenance.

2.3.2 Batch Processing and Analysis

GlobalLink TM Server users must have the ability to run parallel processes and work with more than one file at the same time. Project managers require the ability to import, export, and migrate translation memory assets. Further, for the purposes of budgeting time and costs, users need analysis functionality that will allow them to have a transparent view into a given project's anticipated leveragability. Concrete reporting on the savings generated by leverage (100%, fuzzy, etc.) are needed to help both users and management easily understand the economic benefits derived from a server-based approach to maximizing translation re-use.

2.3.3 Pre-Translation

GlobalLink TM Server users desire the ability to harness the maximum savings potential inherent in translation memory technology, without being dependent upon analysis statistics supplied by their translation resources (be it individual translators or localization service providers). They desire the ability to perform such functions as pre-translating 100% matches prior to routing source materials for translation, thereby further decreasing the cost and time associated with the localization process.

3 System Architecture

When describing GlobalLink TM Server technology as 'next generation,' we are referring to the system architecture. GlobalLink TM Server has effectively abandoned the notion that server-based TM must be 'back-ended' by a database, such as Oracle or SQL Server. Instead, GlobalLink TM Server is based on technology similar to that used in leading search engines. This technology allows massive amounts of data to be processed with performance times unheard of in non-GlobalLink translation memory applications.

3.1 Server Architecture

3.1.1 Leveraging Engine

GlobalLink TM Server's leveraging engine is designed to maximize perfect and fuzzy matches from both memories created in TM Server itself, as well as those exported from other products as TMX and imported into TM Server. It is also possible to modify the segmentation rules used by the system in certain cases where the original TM was created with less standard rules (e.g. when the TM is paragraph based).

3.1.2 Administration (web-based interface, performance, diagnostics, TM creation, import/export, TMX)

To create, delete, or modify a TM in GlobalLink TM Server, users can use the administration tool. TMs can also be imported and exported through this tool. The tool allows admin users to govern who accesses the server TM, and how their entries are classified. It also allows them to run diagnostics and compile performance details.

3.1.3 File Processing (web-based interface, analysis, pre-translation)

GlobalLink TM Server allows batch analysis, translation, and cleanup. These actions are triggered from the user's desktop using a supplied desktop application which allows users to add the files and directories in scope, and then makes the necessary calls to GlobalLink TM Server.

3.2 Client Requirements

On the client side, GlobalLink TM Server gives translation contributors their choice of work environment. Linguists have a choice of three potential desktop systems. They are:

- **GlobalLink Translator Studio for Microsoft Word™**
  GlobalLink Translator Studio for Microsoft Word is designed for individual translators and localization service provider project managers. In order to achieve maximum flexibility in a customer's potential pool of linguist resources, it is Microsoft Word-based. GlobalLink Translation Studio for Microsoft Word has all of the major features consistent with leading computer-aided translation (CAT) tools (Trados, Wordfast, SDLX, Déjà vu). Unlike these tools, GlobalLink Translation Studio for Microsoft Word is not marketed individually to localization services providers, but to end-clients for use with their translation resources.

- **Wordfast**
  With over 13,000 users, Wordfast is the world's second most popular CAT tool. Additionally, it is the only major CAT featuring such benefits as real-time quality assurance, cross-platform compatibility, and the industry's most open, non-proprietary TM format for collaboration.
- **GlobalLink Translator Studio Pro**
  While it is clear that presentation layer technologies based on Microsoft Word provide access to the widest possible array of translator resources, there are several areas in which an editing application is the most appropriate workplace. Specifically, the ability to process tagged files is greatly enhanced by having the workplace rest outside of Microsoft Word. This provides not only greater speed, but also increased ability to prevent errors from being introduced in tagged formats. Translator Studio Pro also incorporates features such as multilingual spell checking, review mode (designed with client reviewers in mind), and a task management tab which is directly linked to instances of GlobalLink Content Director.
- **GlobalLink TXML**
  Users open files for translation from either their desktops, a network, or from a task assigned to them from an instance of GlobalLink Content Director. The main file format used is GlobalLink TXML. A file is either opened directly as TXML or in its native format, and then passes through a set of user-defined parsers to convert the file to TXML before displaying it on-screen.

  Users have the option of choosing what set of parser rules should be applied to the format they are opening, as well as specifying the encoding for the files they are opening.

  The TXML acts as a translation specific wrapper, specifically identifying what is translatable and what is non-translatable in the original file, as well as identifying placables and formatting. The power of TXML is that any file format can be wrapped in TXML based on customizable parser rules. GlobalLink Translator Studio Pro ships with a set of predefined parser rules (filters) for common formats.

3.3 System Architecture Summary

The GlobalLink TM Server system architecture has been designed to allow all translation stakeholders to share enterprise (or centralized) TM assets in real-time over the Internet or a corporate network. Its design allows for speed, usability, and scalability with translation memories of massive size and large numbers of concurrent users.

By architecting the software in such a way as to avoid OEM-ing third party database back-ends, two objectives have been achieved:

- **Time-efficiency.**
  Storing translation memory assets in a specific structure optimized for fuzzy match algorithms generates Google-like performance results for TM searches regardless of the size or complexity of the TM.
- **Cost-efficiency.**
  The GlobalLink TM Server system does not necessitate fees typically associated with the inclusion of third-party technologies as part of the architecture.

Maximum flexibility is achieved through compatibility with all common file types, including all Microsoft Office products (and RTF), and all major tagged file formats, including XML, HTML, SGML, ASP, JSP, and PHP. Flexibility is also achieved by having more than one client-based desktop application that can hook into GlobalLink TM Server.

4 Deployment

As one might expect from any enterprise-wide roll out, successful collaboration between all stakeholers including business users, internal IT resources, and GlobalLink implementation consultants from Translations.com—is critical.

4.1 IT Considerations

As touched on above, some level of due diligence must be conducted with internal IT resources in order to plan the appropriate implementation strategy.

4.1.1 Number of Physical Machines

It may be beneficial to utilize multiple machines for optimal load balance and system performance. Many factors may play a role in considering the best physical location, but chief among them are:

- Number of individual translation memory assets present in the enterprise
- Number of translator resources utilizing the system simultaneously
- Number of administrators, project managers, and/or engineers
- Existing internal IT infrastructure and Internet connectivity

4.2 User Definition, Setup, Management, and Training

Consistent with most technology solutions that approach business problems at the enterprise level, GlobalLink TM Server must maintain certain characteristics and privileges that relate to certain roles within the organization.

4.2.1 System Administrator

The role of System Administrator entails:

- Assigning rights, access, and privileges to various system users
- Configuring the customizable aspects of GlobalLink TM Server
- Designing relevant usage and budgeting report capability
- Managing the migration of legacy translation memory assets
- Ensuring that GlobalLink TM Server drives automation by enabling integration with disparate workflows across the enterprise

Training: 3 to 5 days of hands-on training is recommended for those intending to fulfill the role of System Administrator.

4.2.2 Project Managers and Localization Engineers

Using GlobalLink TM Server, these users typically have responsibility for:

- Translation memory management. Managing the process by which individual translation memory assets are approved for inclusion within an enterprise TM.

- Batch processing and analysis. Managing translation memory assets for maximum re-use. Producing reports on content leveraging, down to the project level if applicable.
- Pre-translation. Prior to assigning work to translation resources, a project manager using GlobalLink TM Server will be able to pre-translate files based on server-side matches and then distribute those items requiring manual translation.
- Management of translation resources. Assigning, monitoring, and managing the work of disbursed translations resources such as external translators, internal translators, or language service providers. Note: with workflow integration within a CMS or GlobalLink Content Director, this task is reduced to monitoring since workflows may be completely automated.

Training: Up to two days of hands-on training are recommended for those intending to fulfill the role of project manager or localization engineer (assuming the individual already has competency performing these tasks in a non-GlobalLink TM Server environment).

4.2.3 Translator or LSP Resources

Actual human translation can be conducted using translators (internal and external) or language service providers. These users can access the system through complete system compatibility with either:

- GlobalLink Translation Studio (Pro or Microsoft Word version), or
- Wordfast (the second most popular translation memory engine)

To the translator or service provider, the desktop application hooks seamlessly onto GlobalLink TM Server, as if it were connecting to a translation memory repository on their local desktop. Therefore, minimal training (if any) is required.

Training: Up to two hours of self-training are recommended for those who have some experience working with any major translation memory engine (Trados, Wordfast, Déjà vu, etc.)

4.3 Migration of Legacy Translation Memory Assets

To begin generating maximum and immediate benefit, the process of importing data and the creation of server-side TMs from legacy translation memory assets should be part of the implementation process. GlobalLink TM Server boasts one of the most flexible and open environments for using TM from many sources, and supports all major formats.

Legacy translation memory assets may be imported into GlobalLink TM Server, as TMX, from all major translation memory engines (including support for all major competitor formats).

The system administrator typically assumes responsibility for managing this process during implementation. Post-implementation, if required, this function can be undertaken by project managers and/or localization engineers.

4.4 Business Process Automation

The system administrator typically assumes the responsibility of integrating GlobalLink TM Server with decentralized content stakeholders (using various workflow and content repositories) across the enterprise. Users may integrate GlobalLink TM Server with a wide variety of authoring, document management, and content management systems:

- EMC's Documentum
- Interwoven's TeamSite
- Fatwire's Content Server
- Percussion's Rhythmyx
- Blast Radius's XMetaL

Additionally, it may be incorporated into systems operating on top of popular database applications:

- Oracle
- Microsoft SQL Server
- Sybase
- MySQL
- IBM DB/2

Finally, GlobalLink TM Server can integrate with a host of other platforms and ERP applications, such as Lotus Domino, SAP, etc.

Some integrations require combining GlobalLink TM Server technology with GlobalLink Content Director, which delivers automated workflows throughout the entire localization process, including change detection, file routing, and status reporting.

5 Conclusions

GlobalLink TM Server presents a robust, reliable, and scalable solution for achieving maximum productivity when faced with the organizational challenge of global content.

5.1 Server-Based Translation Memory

Real-time connection of disparate translators and/or LSP produces immediate significant gains in re-use and the amount of content that can be leveraged. Translators working concurrently can work with incremental efficiencies not possible with standalone client-side translation memory.

5.2 Ownership vs. Hosted Solution

GlobalLink TM Server presents a product aimed at customers with large-scale initiatives who are seeking ownership over their translation memory assets and leveraging benefits, as well as customers seeking to avoid the situation of being forever chained to a service provider who is offering this type of benefit with a "free" hosted solution.

5.3 High Performance Design and Scalability

As the only server-side translation memory solution free of a back-end database, GlobalLink TM Server makes use of Google-like search capabilities to handle searches of translation memories that contain individual translation memory assets numbering in the millions.

GlobalLink TM Server is also architected to accommodate a number of simultaneous users in the hundreds.

5.4 Maintain Flexibility and Integration Capabilities

GlobalLink TM Server users enjoy the ability to integrate seamlessly with over five major content management systems and two different client-based translation memory engines, including Wordfast (widely considered the most powerful translation memory tool for Microsoft Word). GlobalLink Content Director is required for certain integrations and has an aggressive roadmap for future connectors and filters. The GlobalLink Product Suite supports all major file formats.

5.5 Multitask Translation-Related Managerial Activities

Effectively manage the updating and quality of your enterprise translation memory through rights, privileges, and approvals. Batch process and batch analyze files in all major formats. And pre-translate in order to maximize and retain all possible savings associated with leveraging text and re-use. Produce real-time reports which demonstrate cost and time savings from leveragability.

5.6 Support

The GlobalLink TM Server product is accompanied by a dedicated team of support professionals, implementation specialists, and business consultants. Support can be provided 24/7, with tickets submitted by email, via a web-based ticketing system, or over the telephone.

5.7 Small Footprint and Modest Investment

Aside from speed, another advantage of avoiding OEM-ing third-party database technology is also avoiding the fees owed to a third-party upon installation. Further, low required training time, minimal hardware requirements, ease of implementation, and open migration of all major memory formats are all aspects that decrease total cost of ownership and increase return on investment. Lastly, GlobalLink TM Server is based on a per-server license model whereby customers are not charged more based on increased throughput—a unique feature among all common server-side TM technologies