[*Translating and the Computer 21*. Proceedings… 10-11 November 1999 (London: Aslib)]

Case Study:
Document Management and Localization

Dan Dube
Debbie Bucchianeri
Karen O'Neill
Lighthouse Solutions, Inc.
Londonderry, NH, 03063
USA

dan@lighthouse-solutions.com
Phone: +1 603 6274090
Fax: +1 603 627 4060

Introduction

In this paper I would like to address the unique challenges of corporations trying to localize their documentation for a global market and some proven solutions applied to address those challenges. I will talk about the problems, processes and solutions by discussing the case studies of two major corporations that incorporated document management and translation technologies to meet their localization challenges. Lastly, I will describe a general architectural overview of a multilingual document management system, which supports the translation and delivery of information into multiple languages.

Understanding the Problem and the Solution

Cummins Engine Company

THE PROBLEM

Cummins Engine Company is a leading manufacturer of diesel and natural gas engines with 4500 distributors and dealers in 130 countries. They publish over 50 service manuals and 60 parts manuals every year. With this volume of data as well as the need to address global markets, automating the management and localization of information became imperative.

In the past, using traditional flat file systems with no potential for efficient data reuse across publications, information was often inconsistent from publication to publication. Creating a new publication meant duplicating existing files, then relying upon operators to edit and insert formatting elements into the text.

Cummins Engine realized that many of their customers who relied upon their data also required consistency in the information supplied; a requirement that couldn't be met using flat file systems to manage information. In addition to consistency, the publications had to be portable to a variety of software applications and output to multiple media formats. This also was not possible given the proprietary nature of the software products used for editing and publishing information.

To meet the needs of emerging global markets that require the information in local languages, the hard copy of an entire manual was sent to a translator for translation into a target language. Cummins' translation vendor had no way to reuse already translated information; the translation process was costly and the production cycle lengthy.

THE SOLUTION

In the early 1990s, Cummins Engine decided to implement a document management system to manage the reuse of information. They also adopted Standard Generalized Markup Language (SGML) for all data markup to ensure information portability across applications and to extend the range of available output media formats. Combined with the technology afforded by the document management system, information could be chunked into small fragments, allowing for data reuse across publications and easy updates to all publications using those fragments.

In the mid 1990s, Cummins Engine embarked on a project to automate the localization of their publications. They had an urgent need to reduce publication time and translation costs. Their custom solution included custom tools for managing the selection and output of source/target data for translation, the use of translation memory software and post-editing by translators. In addition to the software processes that were implemented, Cummins Engine also adopted standardized wording conventions ("Controlled English") in their authoring process to enable more accurate matching in the translation memory software.

In their system, the source language is English and information is chunked into manageable, reusable fragments. A "diffing" algorithm is applied against the last and current versions of the source language to determine what editing changes have been made. The changed source fragments are run through the translation memory software and a list of unmatched sentences is generated. These unmatched sentences are sent for translation. Upon return, these newly translated sentences are applied to the translation memory software. The source fragment is sent once again through the translation memory software and re-imported into the document management software, in the target language hierarchy.

PROVEN RESULTS

Cummins Engine currently supports Spanish, French, German, Italian, Swedish, and Portuguese languages. The production of localized Operator and Maintenance manuals (400 pages) was reduced from 6-8 months to 10 weeks. They have realized a 65-70% reduction in translation costs primarily due to the use and reuse of common data across publications, the integration with Trades Translation Memory system, and a batch Composition system to automate layout of page formats and graphics, which saved 30% on production costs.

Tweddle Litho Company

Tweddle Litho Company is a full-service technical publishing house dedicated to meeting the needs of the worldwide automotive industry. Tweddle Litho also provides complete data management services including authoring, translation and other related services.

THE PROBLEM

Tweddle Litho Company's mission for their customer (Ford Motor Company) was to produce vehicle owner's literature to support a vehicle release in 30 languages in 60 countries simultaneously. To this end, they embarked on a project to design a new, customer friendly look including more graphics in the layout. Additionally, they developed a culturally neutral global format and determined to meet local requirements for engineering, regulatory, safety and environmental conditions. As an additional goal, they wanted to reduce publication cycle time and costs.

THE SOLUTION

To meet their goals, the authoring process was changed so that authors wrote information based on vehicle. These systems of data were stored as SGML fragments in a document management system. Text was removed from graphics and the new graphics stored in a database. Translated versions of the fragments were stored with the source language for ease of management and ease of use.

With these systems and processes in place, data sharing across owner's literature was optimized and multiple authors could work worldwide from a central repository. In addition, this fragment management allowed for the translation of fragments of data rather than entire manuals. When the translation of these fragments is complete, they can be imported back into the data repository, tracked with their source data.

PROVEN RESULTS

Tweddle Litho currently publishes documents and CDs in over 30 languages. They have realized an increase in shared data to 90%, significantly reducing redundant authoring/translations. They have improved turnaround time for all markets from 6 months to two weeks. They are concurrently managing 4 model years and have been successfully publishing for two years.

A Document Management Model

In this section, I will describe the steps involved in implementing our model document management system, which utilizes SGML as its data markup language and includes processes for translating data from a source language into a target language and stored back into the database. The resultant translated information can be converted from SGML to multiple output formats, including hardcopy composed pages. Acrobat PDF files, and HTML for display in a browser.

Document Management System

Our examples are based on the use of SigmaLink, a document management system by STEP GmbH. SigmaLink is based on client-server architecture. The central application server is developed entirely using JAVA. The server provides the services of the system integrating an SQL-based (Structured Query Language) relational database for central data storage, a structured full-text retrieval system, workflow management software, an SGML-based transformation engine, and an HTTP (Hypertext Transfer Protocol) server.

The client, SigmaLink Workbench, controls access to the information basis. In the Workbench, a user has views of the information objects within the data hierarchy, performs editing tasks with integrated SGML editors (or non-SGML applications), and launches a workflow package that allows complex administrative routines to be automated. In our example, SigmaLink launches Staffware as its integrated workflow package.

Document analysis and DTD development

One of the first steps is to select the initial document to be incorporated in the new system. This document is analyzed for structure and anomalies, and then a DTD (Document Type Definition) is written according to the results of the analysis. Data is converted/tagged with the corresponding elements comprising the DTD. A structured editorial environment is set up to use the DTD and provide dynamic selection of allowed tags.

In our sample environment, the information being managed is a set of User Guides for electronics and small home appliances. A representative sample guide was selected, analyzed, and a DTD written. We used Arbortext Adept Editor and set up the DTD and screen display required with this product.

Hierarchy/data structure

A data hierarchy is created, which allows users a tree navigational view of the data. The hierarchy can be represented in many ways; often the branches of the tree will be created according to categories of publications. The hierarchy is duplicated for each target language.

In our sample environment (see Figure 1), the user would publish 'Care and Use Guides' for various small appliances. There is an English portion of the hierarchy that contains Common Data (information that is useful and useable across publications) and also branches for small appliances, further broken down by appliance type. As shown in Figure 1, there would be separate publications for Blenders, Toaster Ovens and Toasters. Within toasters, Model specific information is represented as separate objects. In this example, Model XXX specific information would include its Features and also Toasting Instructions. Within Common Data, one would find reusable information such as Warranty Information, Performance Guarantee, General Safeguards, and Details for corresponding with the Manufacturer.

Metadata

Each information object, or piece of information stored in the database, has associated with it additional information about the object, or metadata. Some of the metadata is supplied by the system; some is user-defined. Metadata is useful for conducting searches for particular information objects matching defined criteria (e.g. find all the cars with a certain engine type). Metadata can be automatically inserted into metadata property sheets from the SGML elements or their attributes, or can be user-supplied.

Useful metadata for the translation process is to know which target language a source information object is to be translated into. Additionally, useful fields could track word counts, author, and publication title. Figure 2 depicts Metadata fields (Title, Author, Graphics, Target Market and Path) and also shows a sample search criteria in the Toolbar. One may easily search for 'German' in the 'Target Market' Metadata field, and also limit the search to objects that have changed since some user-defined date.

Workflow

The workflow plays a critical part in managing the routing of work in an organization. The workflow specifies tasks to be performed, users allowed to perform them, and the next tasks in the work route. Figure 3 shows a sample workflow for a document review. After proof reading the document, the user decides whether the document is ready to be published or requires post-editing. Once the document is approved, it is then ready for publication. Figure 4 shows the User Interface at the Proof Reading step. The user is prompted with a form and is then able (from the Workflow Manager) to launch a viewer to view the document. The proofreader then decides if the document needs changes (thus determining its route in the workflow) and also can add remarks that will be available for review at the next workflow step.

The user can edit in the Workflow or in the Document Management System. Commands can be performed on route. Scripts can be launched as triggers from one workflow step to the next.

Translation

Source data must be translated into a target language. A step in the workflow could include the 'Translation' step. To accomplish this, various scripts are run. Initially, a script is executed and displays a 'pop-up' box prompting the user to specify the criteria of documents to be translated (i.e. based on the publishing document, date criteria, target language). The appropriate data is exported from the database using external applications. In our example, translation memory and machine translation software is launched on the source object, and then the translated data is re-imported into the database. The target language information object is stored in its appropriate location in the hierarchy. An object-to-object link is created between the source and target object. This link is useful for mapping target objects to their source. When the source object is updated or revised, all target objects linked to that source can be flagged for localization.

In general, the Document Management System manages the information and includes the translation as a workflow step. The applicable data is exported. At this point, the user could do a number of various options:

- run the files through in-house translation memory and/or machine translation tools, and then send the results to the translation vendor for post-editing and cleanup
- include the translation vendor in the workflow and allow them to directly manipulate the exported information and check-in localized files back to the document management system
- send the exported files to the translation vendor for localization, and import the localized files upon completion.

Publishing

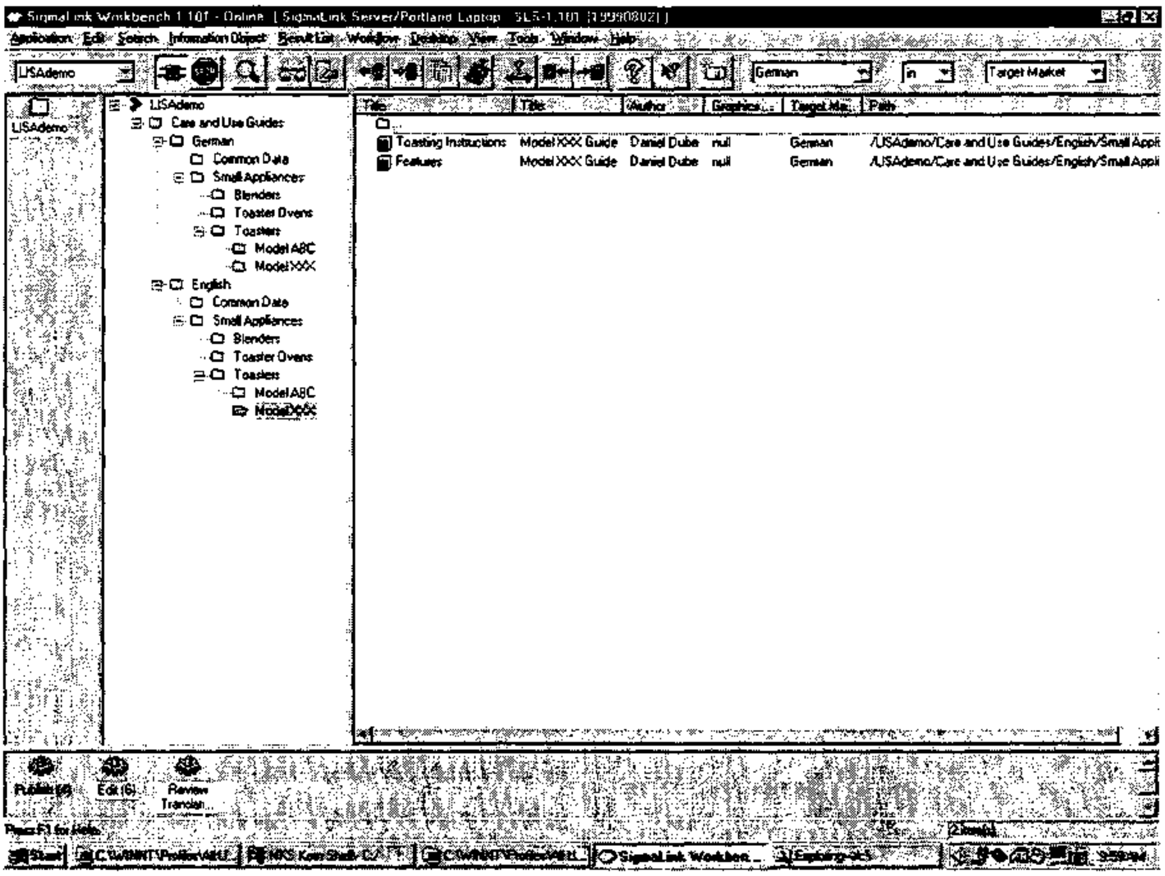
In our sample environment, the document is published by the execution of scripts that export the data, convert the SGML to HTML and then launch a browser to view the document. However, the system can be configured to convert the SGML data to any format and output (CD, web, hard-copy) for any number of documents.

Conclusion

The case studies of Cummins Engine and Tweddle Litho Company have demonstrated the proven results of migration to SGML for information authoring and storage, reuse of SGML information components, and workflow automation of the localization processes. Since the implementation of the automated localization environment at these customer sites, both companies have experienced the following dramatic results:

- Significantly lower translation and localization costs
- Dramatically improved time-to-market for localized information products
- More consistent information, leading to a higher quality localized product

Figure 1 - Sample Hierarchy

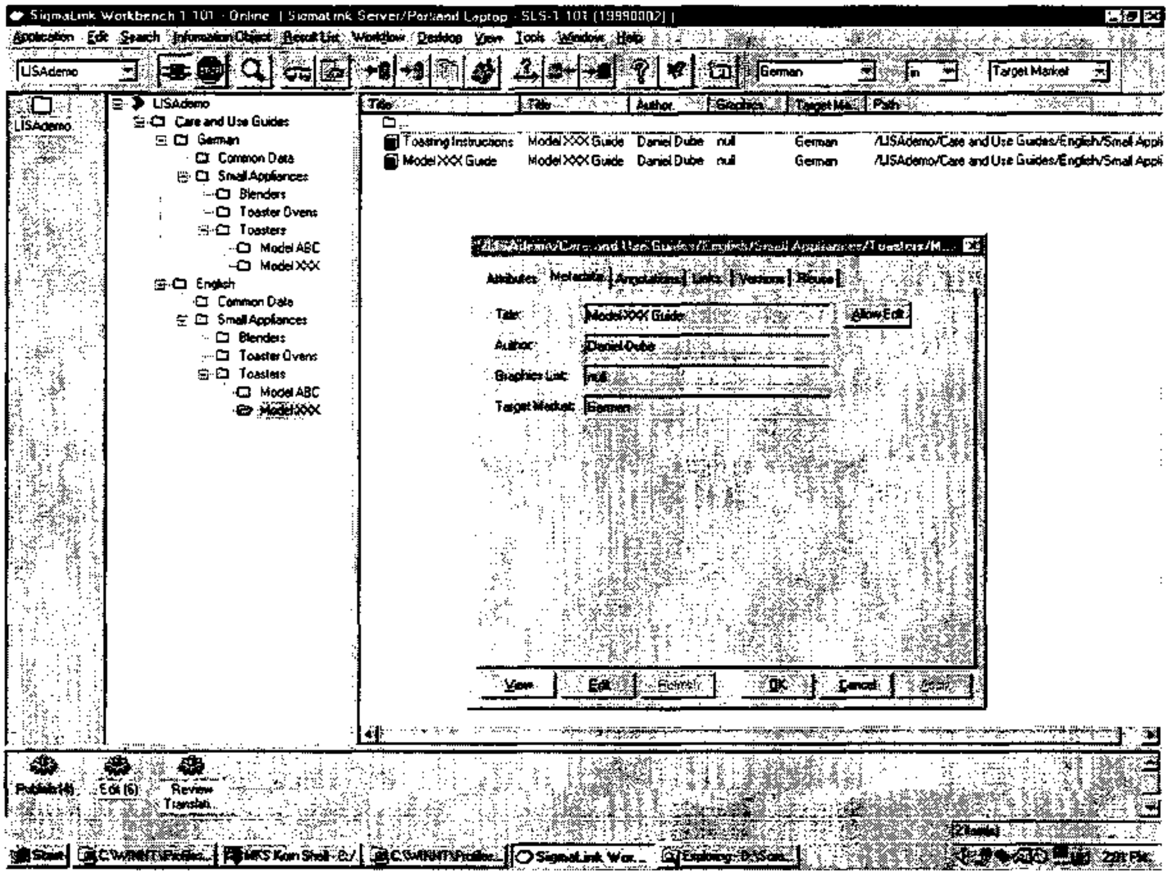


SignalLink Workbench 1.101 - Online [SignalLink Server/Portland Laptop SLS-1.101 (19990802)]
Application Edit Search Information Object Result List Workflow Desktop View Tools Window Help
USAdemo
German
in
Target Market
USAdemo
Publish (4)
Edit (6)
Review Translati...
Press F1 for Help
2 item(s)
Start

- USAdemo
  - Care and Use Guides
    - German
      - Common Data
      - Small Appliances
        - Blenders
        - Toaster Ovens
        - Toasters
          - Model ABC
          - Model XXX
    - English
      - Common Data
      - Small Appliances
        - Blenders
        - Toaster Ovens
        - Toasters
          - Model ABC
          - Model XXX

Title	Title	Author	Graphics...	Target Ma...	Path
..					
Toasting Instructions	Model XXX Guide	Daniel Dube	null	German	/USAdemo/Care and Use Guides/English/Small Appli
Features	Model XXX Guide	Daniel Dube	null	German	/USAdemo/Care and Use Guides/English/Small Appli

Figure 2 - Sample Metadata



Title	Title	Author	Graphics	Target Ma...	Path
..					
Toasting Instructions	Model XXX Guide	Daniel Dube	null	German	/LISAdemo/Case and Use Guides/English/Small Appli
Model XXX Guide	Model XXX Guide	Daniel Dube	null	German	/LISAdemo/Case and Use Guides/English/Small Appli

/LISAdemo/Care and Use Guides/English/Small Appliances/Toasters/M...

Title: Model XXX Guide

Author: Daniel Dube

Graphics List: null

Target Market: German

Figure 3 - Staffware Workflow Procedure

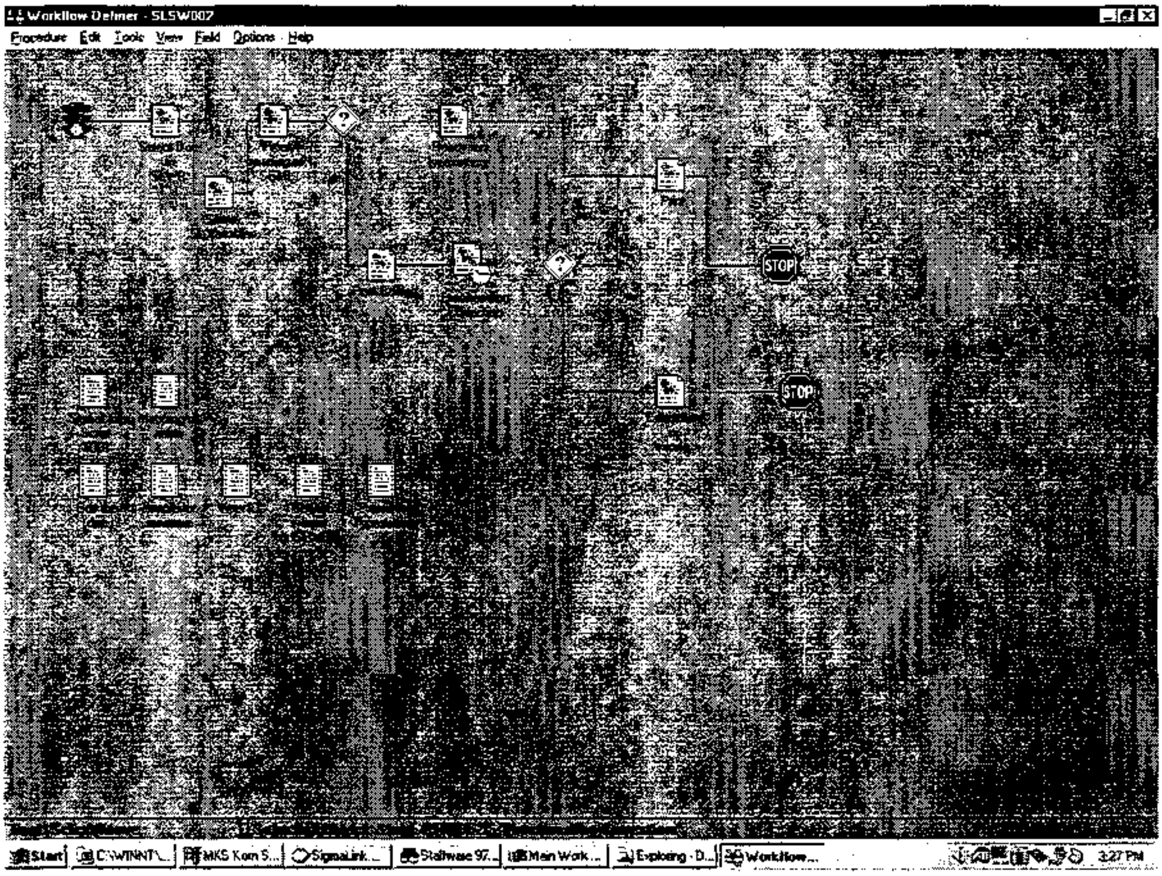
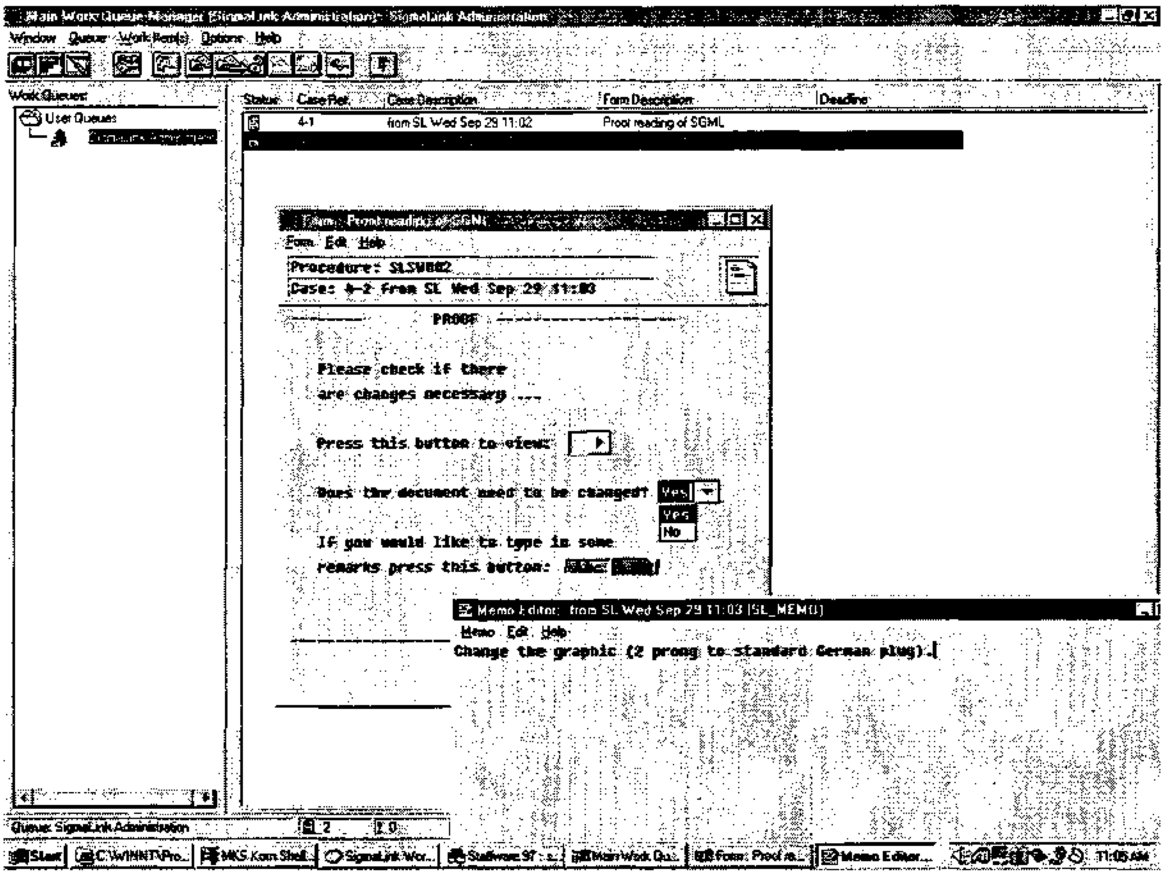


Figure 4 - Staffware Work Queue Manager: Proof Reading Step


Main Work Queue Manager (SignalLink Administration): SignalLink Administration
Window Queue Work Item(s) Options Help
Work Queues:
User Queues
Form: Proof reading of SGML
Form Edit Help
Memo Editor: from SL Wed Sep 29 11:03 [SL_MEMO]
Memo Edit Help
Start
C:\WINNT\Pro...
MKS Korn Shell...
SignalLink Wor...
Staffware 97 ...
Main Work Qu...
Form: Proof re...
Memo Editor...
11:05 AM

Status	Case Ref.	Case Description	Form Description	Deadline
	4-1	from SL Wed Sep 29 11:02	Proof reading of SGML	

Procedure: SLSWB02

Case: 4-2 from SL Wed Sep 29 11:03

PROOF

Please check if there
are changes necessary ...

Press this button to view: ▶

Does the document need to be changed? Yes ▾
Yes
No

If you would like to type in some
remarks press this button:

Change the graphic (2 prong to standard German plug).

Queue: SignalLink Administration 2 0