A Multimodal Predictive-Interactive Application for Computer Assisted Transcription and Translation

Vicent Alabau, Daniel Ortiz, Verónica Romero, Jorge Ocampo
Institut Tecnològic d'Informàtica
Universitat Politècnica de València
Camí de Vera, s/n
Valencia, Spain
{valabau,dortiz,vromero,jocampo}@iti.upv.es

ABSTRACT

Traditionally, Natural Language Processing (NLP) technologies have mainly focused on full automation. However, full automation often proves unnatural in many applications, where technology is expected to assist rather than replace the human agents.

In consequence, Multimodal Interactive (MI) technologies have emerged. On the one hand, the user interactively cooperates with the system to improve system accuracy. On the other hand, multimodality improves system ergonomics.

In this paper, we present an application that implements such MI technologies. First, we have designed an Application Programming Interface (API), featuring a client-server framework, to deal with most common NLP MI tasks. Second, we have developed a generic client application. The resulting client-server architecture has been successfully tested with two different NLP problems: transcription of text images and translation of texts.

Categories and Subject Descriptors: I.5 [Pattern Recognition]: Miscellaneous

General Terms: Design

Keywords: Handwritten recognition, Machine translation, Interactive framework, Multimodality

1. INTRODUCTION

Traditionally, Natural Language Processing (NLP) applications have focused on fully automatic systems. However, since their performance is far from being perfect, automatic systems cannot replace the experts. Typically, human experts use automatic systems as follows: first, the system generates its output in a fully automatic way; and second, the human expert revises this output in order to achieve standard quality results. Such post-edition solution is rather inefficient and uncomfortable for the user.

As an alternative to post-editing, Interactive and Multimodal Interactive (MI) approaches have been proposed recently [1, 5]. The user feedback is provided by means of pen strokes on a touchscreen and/or more traditional keyboard and mouse operation. In this approach, the automatic system and the human expert both cooperate to generate the final solution. The user feedback allows to improve the system accuracy, while multimodality increases system ergonomics and user acceptability. MI is approached in such a way that both the main and the feedback data streams help each other to optimize the overall performance and usability.

Copyright is held by the author/owner(s).
ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.
ACM 978-1-60558-772-1/09/11.

2. MI APPROACH

In the MI approach, the user is involved in the transcription process since she is responsible of validating and/or correcting the system output. The protocol that rules this process, is formulated in the following steps:

- The system proposes a fully automatic hypothesis
- The user validates the longest prefix which is error-free and enters some pen strokes and/or some keystrokes to correct the first error in the suffix
- If pen strokes are available, an *on-line* HTR feedback subsystem is used to decode this input
- A new extended prefix is produced based on the previous validated prefix, the on-line decoding word, and the keystrokes. Using this new prefix, the system suggests a suitable continuation of it
- These previous steps are iterated until a perfect result is obtained

3. DEMO OVERVIEW

In this work, we present an application that implements the previous approach in a generic way. The application has been successfully tested with two different NLP problems: Handwritten Text Recognition (HTR) and Machine Translation (MT).

First, based on the MI protocol, we extracted a generic subset of primitives for most common NLP tasks, and designed a client-server Application Programming Interface (API) that allows client and server applications to communicate through sockets.

Three basic functions summarize the API:

set_source : selects the source phrase to be transcribed or translated

set_prefix : sets the longest error free prefix and amends the first error with the keyboard

set_prefix_online : sets the longest error free prefix and amends the first error with pen strokes

Next, a generic client application was developed to use the API. The client application (Figure 1) is responsible for showing the user interface and capturing the user actions

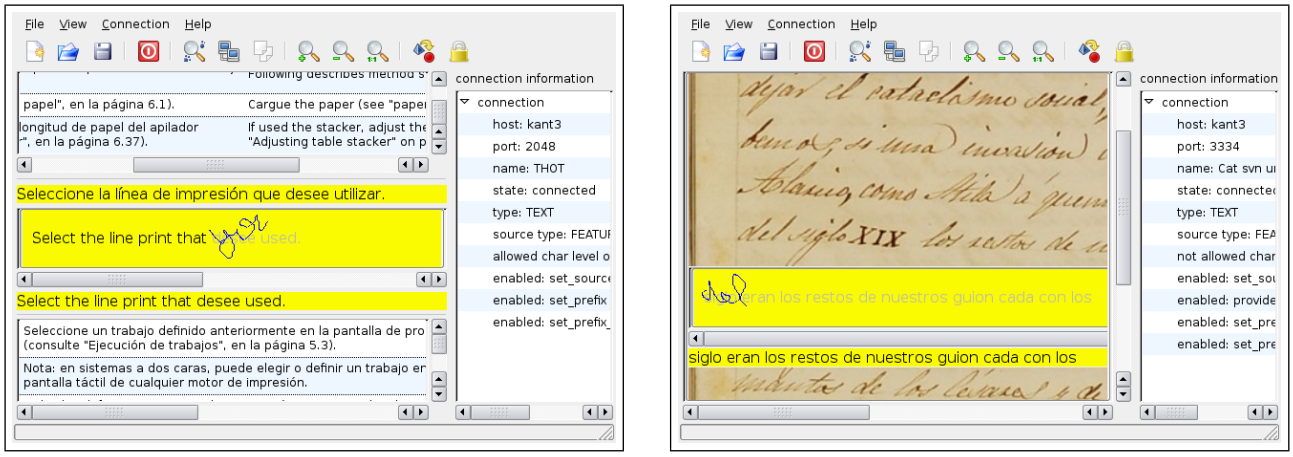

Figure 1: Left: illustration of the client application with the Machine Translation view. Right: illustration of the client application with the Handwritten Text Recognition view. Note, in both views, the input boxes in yellow background. The one in the bottom is used to interact with the keyboard. The big one serves as canvas for pen touchscreen interaction. The right panel gives information about the server the user is connected to.

on the different modalities of interaction, i.e. keyboard and pen strokes. The most important components of the client application are the visualization area, the keyboard input box and the pen input box. The visualization area, which uses **set_source** to select the source phrase, depends heavily on the task. Thus, specific visualizations must be implemented for each task. On the contrary, the keyboard (**set_prefix**) and pen (**set_prefix_online**) input boxes can be reused without any modification. The former captures the input from the keyboard, which is assumed to be error free. The latter captures pen strokes as a series of points.

Finally, servers for MT and HTR were built. Servers combine all the information received from the client and compute a suitable solution. All the decoding process is performed at the server side. For the MT task, a server implementation of the Thot decoder was used [3]. The online and offline HTR systems are based on Hidden Markov Models (HMM) and bigram language models. However, the preprocessing and feature extraction techniques differ between them [5].

4. PREVIOUS WORK

Although the demo presented in this paper shares a common MI framework to the ones presented in [4] for HTR and in [2] for MT, it has some neat advantages over them.

First, the demo in [4] is a web based demo. As such, internet connection is needed. The demo presented in this paper can be used either with or without internet connexion. Besides, the web based demo poses difficulties to other interaction modalities, such as speech, which, up to now, cannot be nicely integrated in the web.

Second, regarding the MT demo in [2], our demo allows pen touchscreen interaction, in addition to keyboard and mouse. Furthermore, thanks to the client-server architecture, specialized MT servers in mainframe computers can be run for complex tasks, if needed.

Finally, and most importantly, the demo shown in this paper serves as a test bed for new interaction modalities such as speech, or to experiment with other NLP problems, namely speech transcription or text prediction. In addition, it allows the design of different user interfaces (or views) to approach same NLP problem in a more ergonomic and comfortable way.

Acknowledgment

This work has been supported by the EC (FEDER), the Spanish MEC under grant TI N2006-15694-C02-01, the research programme Consolider Ingenio 2010 MIPRCV (CSD2007-00018) and the UPV (FPI fellowship 2006-04)

5. REFERENCES

[1] S. Barrachina, O. Bender, F. Casacuberta, J. Civera, E. Cubel, S. Khadivi, A. L. H. Ney, J. Tomás, and E. Vidal. Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1):3–28, 2009.

[2] A. Lagarda, L. Rodríguez, E. Cubel, E. Vidal, and F. Casacuberta. Transtype 2. un sistema de ayuda a la traducción. In *Proceeding of the SEPLN: XIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje*, pages 345–346, Septiembre 2003.

[3] D. Ortiz-Martínez, I. García-Varea, and F. Casacuberta. Thot: a toolkit to train phrase-based statistical translation models. In *Tenth Machine Translation Summit*. AAMT, Phuket, Thailand, September 2005.

[4] V. Romero, L. A. Levia, A. H. Toselli, and E. Vidal. Interactive multimodal transcription of text image using a web-based demo system. In *Procedings of the International Conference on Intelligent User Interfaces*, pages 477–478. Sanibel Island, Florida, February 2009.

[5] A. H. Toselli, V. Romero, and E. Vidal. Computer assisted transcription of text images and multimodal interaction. In *Proceedings of the 5th Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, volume 5237 of *LNCS*, pages 296–308. Utrecht, The Netherlands, September 2008.