Multilingual Information Access: the MuST System and its Integration with Systran

Chin-Yew Lin and Eduard Hovy USC/Information Sciences Institute Marina del Rey, CA 90292-6695, USA Tel: 1-310-822-1511 E-mail: {cyl,hovy}@isi.edu

10

1. INTRODUCTION

The goal of the MuST [1] project is to develop a prototype system to facilitate not only retrieving documents from multilingual collections, but also to summarize and translate the retrieved document into the user's preferred language. We focus on the integration of state-of-the-art technologies, try to identify the critical path of enabling multilingual information access, and propose possible solutions. As far as possible, the system employs existing resources and products, such as the search technologies from America Online (AOL)/Personal Library System (PLS) [2], and online Internet search engines. It incorporates web spider technology enabling users to target their areas and languages of interest. It provides multilingual summarization technology developed at ISI [3] enabling users to quickly judge the relevance of the retrieved documents. It also integrates deep and shallow translation engines, both built at ISI [4] and commercially available, for online browsing of foreign language texts. We use the World Wide Web as our multilingual document sources and assume English is the source language. MuST can handle the languages English, Arabic, Japanese, Spanish, and Bahasa Indonesia. We plan to add more languages in the near future.

In this demonstration, we show MuST in the context of retrieving, summarizing, and translating Indonesian text.

	the <u>subtrance (UEL)</u> of a Web page here	5 S.
http://abcuwes.gi.ium/s atkend.biml	eestions/living/infourment/	a111070 -
		4
Ownel		_
Statement Strengther Constitutions	3	
Translate trans English to Spec	nish 📕 using Op Trensferor	-
10.00	2	
Heder		TION
C Semenative C Translate	¹⁷ Summarize and Translate	
C Summarije C Translate	¹⁷ Semanation and Translate <u>1996</u> g and translate year private document	a in seconds.
C Summarije C Translate		
C Summarije C Translate		s in seconds.

Figure 1. Entry page of MuST integrated with Systran online translation.

In order to demonstrate the multilingual processing capability and flexibility of SUMMARIST embedded in MuST, Figures 1 and 2 present the integration of SUMMARIST with the online machine translation engine built by Systran Software Inc., a market leader. Figure 1 shows the entry page. Users can type in free text or a specific url, enter query words, choose the number of summary sentences, select the source and target language pair, set the preferred online translation engine, and specify which action to perform by the service (summary only, translation only, or both). Figure 2 shows the result of the summarized and translated page (English to Spanish) of the news about the "saneness of the Unabomber" that appears at ABCNews site on October 13, 1999.

ABCNEWS.com : How Sana is Ted Kaczynski? Summary		
(vi) A post-to-to-	rð senglingad præktskelst pærkonstall a smalt ærdanud evaluation af Teil Kanop sendi illumarni, inskuling arki rephræsia.	
(W) Destingation	Ted Kauryohi seys he's not insare.	
INCOME AND ADDRESS OF	white, payedicately expects say, that functional believes he's eventally well.	
	he Kind of behavior you'd sepect from a person with achicophrania, which was person Raciyouth offsechie around in the Unaderator case.	
professional and	samp people with existing/menia, the thir tracty and, who discourse with the mediate captures, " tays fragment flow, discourse of even people periods of the transmit and all Mediates and same of these with events of tracy with infer the press.	
(H) the stepheters meaned in the result (H) the code carry (SI) the code carry later (SI) the code carry later (SI) the code carry many transition of the state carries do the state of the state of the protocological state	In tool acceptedation and the second	
Etraboritier %		

Figure 2. Summarized and translated page of the news about "the saneness of the Unabomber."

- C.-Y. Lin, "Machine Translation for Information Access across the Language Barrier: the MuST System," presented at Machine Translation Summit VII, Singapore, 1999.
- [2] AOL, "PLWeb Turbo," 3.0 ed: America Online Inc., 1998.
- [3] E. Hovy and C.-Y. Lin, "Automating Text Summarization in SUMMARIST," in Advences in Automatic Text Summarization, I. Mani and M. Marybury, Eds. Cambridge: MIT Press, 1999, pp. 81– 94.
- [4] K. Knight, "Machine Translation of Natural Languages," Information Sciences Institute, University of Southern California, 1999.