System Demonstration

InfoRaptor - SYSTRAN'S Multilingual Data Retrieval System

Denis A. Gachot

SYSTRAN Software, Inc.
7855 Fay Avenue, Suite 300
La Jolla, CA 92037
dgachot@systransoft.com
Tel: 619-459-6700

Introduction:

For over 16 years, SYSTRAN'S linguistic department has been using a parser-based data retrieval tool to collect data exhibiting particular grammatical or semantic features for use in development of machine translation systems. Today, InfoRaptor, a cross-lingual information retrieval system, utilizes SYSTRAN'S technology to query and retrieve sentences from a pre-parsed text.

System Builder and Contact:

SYSTRAN Software, Inc.	
Project Manager & Director of Software Engineering	Jin Yang - jyang@systransoft.com
Director of Sales & Marketing	Reba Rosenbluth - rrosenbl@systransoft.com

System Category:

InfoRaptor is under development for clients in the US Government, and until the pilot release, is available only to development sponsors. The pilot system and other subsequent releases will be made commercially available.

Alpha system release:	November 1997
Commercial pilot release:	To be announced
All-encoding-schemes release:	To be announced

System Requirements:

IBM compatible PC - 486 or faster processor
SVGA + monitor
16-200 MB free disk space
16-20 MB RAM
Windows 95 or Windows NT

System Description:

InfoRaptor is a cross-lingual information retrieval system based on SYSTRAN parsing and machine translation technology.

InfoRaptor integrates machine translation technology at all levels of information retrieval. It consists of two parts: a document pre-processor and a browser. The pre-processor, which is a SYSTRAN MT system, parses and translates documents, resulting in a searchable file which contains a wealth of information on the morphological, syntactic, and semantic level, as well as information on the output translation. From the pre-processed documents, the browser performs by structured queries which allow great flexibility to describe sophisticated linguistic information at various level, or free-form queries which doesn't require any expertise. This integration spans the uses of SYSTRAN'S large multilingual dictionaries, its fine-grained parsing capability, its semantic disambiguation tools, and actual translation. Multilingual retrieval is realized by allowing a foreign language text to be queried in its original language and/or in English or in an abstract language-independent form, and by supplying the user with the results of the search in both the language of the document and in the user's language. The languages for which the browser will be available are determined by the language pair translation systems in production. Their number is constantly growing.