

6.2 Shakti : A kit for building a Retargetable Machine Translation System

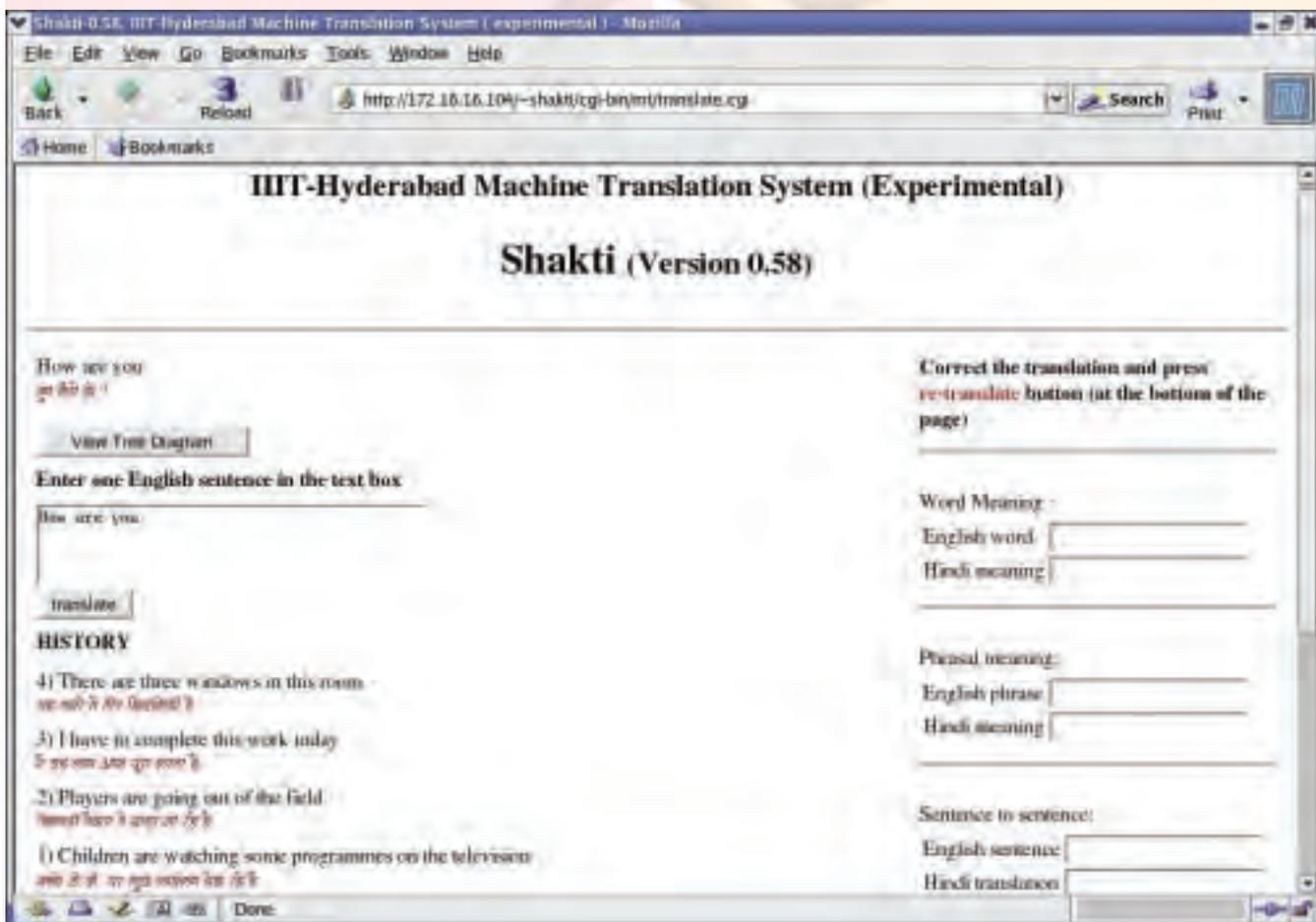
Shakti Machine Translation Kit is a do-it-yourself kit for building an MT system from English to any Indian language. It has been created based on Shakti Machine Translation system being developed at Language Technologies Research Centre, IIT, Hyderabad.

Shakti machine translation system has been designed to produce machine translation systems for new languages rapidly. It has been already developed for from English to three different Indian languages – Hindi, Marathi and Telugu. The System is so designed that many of the benefits of improvement to the system flow automatically to outputs in all the languages.

The Shakti machine translation system is also designed to take ready made sub-systems either as black boxes or as open source software and incorporate them into itself. The simplicity of the overall architecture makes it easy to do so. Available English analysis packages have been extensively adapted by the Shakti machine translation system.

A number of system organization principles have been used which have led to the rapid development of the system. While the principles by themselves might not appear to be new, their application to machine translation in this manner is unique.

The architecture of Shakti is highly modular. The complex problem of MT has been broken into smaller sub-problems. Every sub-problem is a task which is handled by an independent module. The modules are put together using a common extensive representation using trees and features. Modules are pipelined and the output of the previous module becomes the input of the following module. The overall system architecture is kept extremely simple. All modules operate on a stream of data whose format is fixed. They read the data in the Shakti standard format (SSF) and generate output in the same format. Thus, even though the format is fixed, it is extensible in terms of attributes or analyses.



This approach follows the dictum: "Simplify globally, complicate locally." However, since the number of modules is large and each local module does a small job, the local complexity remains under tight control for most of the modules.

Experience has shown that this methodology has made debugging as well as the development of the system convenient for programmers and linguists alike. In case, an output is not as expected, one can quickly find out what went wrong (that is, which module did not function as expected). In fact, linguists are using the system quite effectively to debug their linguistic data with ease.

Shakti system combines rule-based approach with statistical approach. The representation (SSF) is designed to keep both kinds of information. Although the system accommodates multiple approaches, it has a strong bias for linguistic analysis which serves as the backbone. Statistical and other approaches are interpreted in linguistic terms wherever possible. As a result of this, generalizations can be made which are not possible otherwise. Only when no linguistic interpretation is possible, the information is kept as it is. Thus, although the system has a strong linguistic bias, it is capable of dealing with non-linguistic information or patterns in their own right.

Some sample outputs from the Shakti system are:

I have eaten such delicious apples earlier also — >

मैं ऐसा स्वादिष्ट सेब पहले भी खा चुका हूँ

Children are watching some programmes on television in the house — >

बच्चे मकान में टीवी. पर कुछ कार्यक्रम देख रहे हैं

I saw the bird on the tree in the garden near the lake — >

मैंने झील के पास उद्यान में पेड़ पर चिड़िया देखी

We are all thieves and they are all liars. — >

हम सभी चोर हैं और वे सभी झूठे हैं

One may try translating some sentences from English to Hindi online.

The URL of the Shakti machine translation site is:

<http://216.236.98.137/~shakti/>.

The Shakti-kit is being released very soon. Anybody interested in trying their hands on the kit, can send a mail at the following e-mail address: shakti_kit@iiit.net.

*(Courtesy: Rajni Moona (r_moona@iiit.net)/
Dipti M Sharma (dipti@iiit.net) LTRC,
International Institute of Information
Technology,
Gachibowli, Hyderabad
Phone: 040-23001412
Fax: 040-23001413*