

# Multilingual Information Service System for the Beijing Olympics, a playground of MT technologies

Wei quan LIU

Beijing Network and Multimedia Laboratory  
6<sup>th</sup> JBXXG, No. 3 Yumin East Rd.  
Beijing, China, 100029  
liuwei quan@capinfo.com.cn

## Abstract

This paper introduces the Multilingual Information Service System being implemented for the Beijing Olympics. Multilingual machine translation is an important component in this system. This real world application asks for advanced as well as mature and proven technologies, where MT is challenged. However, by appropriately choosing domain and scenario, current MT technologies are successfully integrated in the pilot system. Future applications ask MT to have better performance in readability, lexicon coverage and more efficient. Multilinguality support and fast language adaptation is highly desired by such real world systems.

## 1 Introduction

The XXIX Olympic Games is going to be held in Beijing, 2008. Both the municipal and the Beijing Olympic Organizing Committee (BOCOG) are working intensively to make sure this game is a successful and outstanding one. In the 2nd half of 2001, shortly after Beijing won the rights to host this game, they issued the “Action Plan for the Olympic Games”, in which the difficulties in languages are addressed.

In this action plan, realizing the language barrier is a major issue to hinder people from effective communication, the organizers promise “...to ensure that any body at any time and in any place can enjoy the information service that is affordable, comprehensive, personalized and multilingual...”

It's estimated that 600 thousands foreign spectators will come to Beijing during the game time, they could come from any nation and speak many kinds of languages. Though Chinese, English and French are appointed as the working languages, but in terms of more than 120 national delegations, it's not enough to cover all their native tongues. People will meet difficulties when acquiring and exchanging information using different languages. Especially in China, where Mandarin is

predominant, English speaking people is rare, let alone other foreign languages.

Since 2002, supported by the Chinese Ministry of Science and Technology, the Beijing Municipal government and BOCOG, partners from industry and academic conducted a serial of pilot studies on developing a multilingual information service system for the Olympics. The goals are to use the latest, as well as mature and proven technologies to help the Olympic participants overcome the language problem.

Encouraged by the achievements of these research activities, BOCOG initiated the “Multilingual Information Service System” project in April 2005. The system will provide Olympic participants multilingual information on both the city and the games, in order to eliminate their dread feeling caused by unknown languages and help them enjoy the games better.

This paper serves to introduce the previous work done and the system kicked off by BOCOG which is still in planning phase when this paper is written.

## 2 The requirements and functional specifications

BOCOG organized the user requirement analysis as the first step towards implementing the system. The focus is on the following issues: user population distribution, the languages to cover, the catalog of information that the system is going to provide, and the means of how to provide the information.

The analyzing procedure includes:

- Investigating the requirements from all the functional divisions within BOCOG. Inquiry forms and seminars were designed as tools to clarify their responsibilities during the game, their target users and the service specification
- Probing the needs from the target user. Marketing analysis methodologies were used to explore the needs from typical users
- Identifying the catalog of information and the potential sources of each kind of

information. The identified sources include government agencies, commercial entities and BOCOG's internal systems

- Determining the means of providing these services, e.g. through a website or a call center. The choice is decided by the preference of the users, the nature of each service and many other factors. The practice of former Olympics and the situation in Beijing municipal should also be considered

The requirement analysis is described in more detail by the following subsections:

## 2.1 User population and languages to cover

Based on the analysis of former Olympic Games and the estimation data for Beijing Olympics, the spectators of Olympics and Paralympics will be between 7 to 10 million, the registered members will be around 300 thousands.

User groups	Features
Spectators	Total 7-10 millions anticipated
- Beijing citizen	6 millions
- Domestic from other cities	1.2 millions. They need information in Chinese on game related and city related e.g. lodging, traffic and touring
- Foreigners	600 thousands. They need information in multiple languages on the above and other categories, like everyday knowledge of how to exchange currency
Registered members	Total 300 thousands
- Olympic Family	55 thousands, including International Olympic Committee and National Olympic Committee officials, national delegations etc. They need intensive game information in more languages other than the official ones
- Media persons that authorized	25 thousands. They need real-time dissemination of the competition results, in their preferred languages
- Sponsors and market partners	110 thousands. They need both city and game related information. Intensive requirements on location

	based information
- Paid staff - Volunteers	5 thousands and 80-100 thousands respectively. They need direct language translation support when serving others

Table 1: The user population configuration

For the languages that this system is going to cover, the mandatory ones are Chinese (Mandarin and Cantonese), English and French. To benefit more people, there should be other widely spoken languages like Spanish. Since more spectators come from surrounding countries like Japan and Korea, their languages should be included.

The final decision is waiting for the ticket selling program being made. Then the distribution of the languages spoken by spectators could be more precisely estimated.

## 2.2 Information catalog

The information that this system is going to provide comes from two major categories: city and the games. The content is going to be supplied in multiple languages.

City information includes:

- Information for tourists: lodging, site seeing, shopping, entertainment, restaurant and transportation
- Basic information and common sense on weather, currency exchange, emergency numbers, medical care services
- Location based information, by combining the geographical information with dynamic content from the city, such as the guiding introduction to a tour site, or the description and location of a restaurant etc.

Game related information includes:

- Ticket selling information, where and how to buy tickets
- The detailed schedule of the games
- The introduction to the athletes and teams from every nations and for each competition
- The description of the venues, their locations, the means of transport to get there, what needs to know for entering and watching a game
- Real time release of the competition results and the medal tally
- The basic knowledge of the sports, such as competition rules
- The schedule of various Olympic thematic events
- The schedule of game arrangement in associate cities, such as the equestrian in Hong Kong

### 2.3 Defining the multilingual information services

The system will integrate all these kinds of information and provide them to users with appropriate means.

A comprehensive website will be built to serve all the users before and during the game time. Before the game, it will publish information on ticket selling and the game schedule. Authorized hotel reservation service will be linked to this site. It helps people to order tickets and book hotels when needed.

At game time, this website will deliver useful city information and real time game information.

It supports WAP access, so mobile phone users can visit it on the move. It will also push information to the existing kiosk system in the city. There will be more than 3,000 kiosks available in year 2008.

Mobile phone users can subscribe short message (SMS) or multimedia message (MMS) services from this system. The system will push information in the preferred language based on the user profile. SMS and MMS are also going to be used in pair, in the online phrase book translation service. For instance, an English speaking user may send "*How much is this?*" as a text short message to the service number. Then shortly after, the service sends back its Chinese translation and the pronunciation in a MMS, which could be replayed in front of a shop keeper.

The front-end devices supported by this system include customized handheld. They will provide multilingual location based services. The functions designed for spectators and registered members are different. To spectators, the location based information of tour sites, shopping centers, restaurants, subway stations etc. are preferred. To registered members, especially the sponsor's employees, their favorite information is about the venues and functional centers, where their work is conducted.

A multilingual call center will be built as the most important part of this system. Its size is larger than any single one that can be found in Beijing. Direct human voice interaction is regarded as the principal means of multilingual service. There are five types of services:

- Language assistant service. There are professionals on specific languages. They behave as interpreters. A user, for example a volunteer worker, may not understand the language spoken by a person that she wants to help. Then she can call an appropriate language assistant to translate their conversation over the phone.

- Transport advising service. It helps people who ask about how to take bus or subway and make connections. Since Beijing has complicated public transportation systems, visitors can easily get confused. During game time, taxis will be restricted to approach all the venues because there are large pedestrian zones for the crowd. Spectators are suggested to take bus or subway to go to venues from their hotels.
- Hotel reservation service. Together with the comprehensive website, this call center service is going to help more than 1 million spectators to find hotels.
- Game information inquiry. During game time, this service is going to provide real time news from the games. This service largely depends on multilingual IVR technologies, so human agents are responsible to updating the multilingual content, instead of answering the call.
- Emergency helpdesk service. When people encounter emergent circumstance, it's advised to call this unique number for help. It is designed especially for foreigners, who know little about how to access different city functions. In China, the emergent medical service, fire department and police each have their own numbers. They usually speak Chinese or limited English. The multilingual call center for the Olympics will bridge them and help foreigners by speaking their familiar languages.

### 3 The pilot systems

Pilot systems were being continuously developed in order to investigate how the latest language technologies could be utilized in real world applications for the Olympics, their applicability and how they should be improved. These technologies mainly come from research partners supported by research programs. Using their multilingual machine translation, speech recognition and text-to-speech components, three major pilot systems have been built until now.

The first pilot works with the existing kiosk system in Beijing. It provides travel information on these kiosks in three languages: Chinese, English and Japanese. Its purpose is to let tourists to get weather, tour site description and transportation information conveniently. This is designed as a speech enabled application. Since there are microphone and speaker on a kiosk, users could navigate the pages by using speech commands. The text displayed subsequently could be read by TTS. Users welcome this application because it's

easy to access and the information is updated in real time.

The second pilot is not for end users. It uses the translation memory technologies at the backend of the BOCOG's websites. These webs are publishing in English, Chinese and French. The original content is usually in Chinese, so it's the editors' responsibility to do the translation. This work is time limited, since the three language versions should be synchronized. It's a burden to the human translators to keep efficient while maintain the quality of the output.

Chinese to English and Chinese to French translation memory technologies were introduced to them. A user-friendly interface is provided, so human translators work with their familiar word processor software, TM works at the background. This technology supports group working. Members in a group share the same memory database. With these help, both the consistency and the speed got improved greatly.

The third pilot is a speech enabled multilingual application on handheld devices. Currently it supports Chinese and English. It provides location based information about the major tour sites in Beijing. English to Chinese bidirectional MT is embedded on this device. Together with an indexed phrase book, it becomes a pocket translator which proves to be useful to foreigners. This device is being tested among selected foreign visitors. They can locate and browse the description to a nearby tour site, or use the translator to communicate with taxi drivers simply.

Speech translation technology was once integrated on this device. But due to its performance and usability, it seems not applicable under the real world noisy environment and user's spontaneous speech.

#### **4 The application of multilingual MT**

In the pilot systems and real Olympics system that is planned to have, MT is or going to be widely applied. In the coursing of applying MT technologies, there are characteristics and guiding rules worth to be noticed.

"Multilingual" MT is required in most of the applications. There are three official languages and others are to be added in the future. The information services are expected to be provided in all languages equally. The information source is in Chinese in most of the cases, so these services are heavily depended on MT that translating from Chinese to other languages.

MT is integrated with other language technologies, such as speech recognition and TTS. They can work collaboratively if the intermediate result of one component can be used by other ones.

The lattice generated by speech decoder may help MT in parsing. The intermediate result when MT is generating sentences may help TTS to do the text preprocessing and prosody analysis.

The information services are implemented on different hardware and software platforms, from telephone to handheld devices. MT components should be tailorable and the kernel is platform independent. Currently it's still difficult to implement a full bloom multilingual MT engine on a handheld, mainly because of the limited memory.

In some services, the MT is not applied to face end users directly. Language professionals are using it in stead. They expect the MT engine has better lexicon coverage and readability. The machine generated result is not just carrying the meaning but also could be used as a good foundation for further human improvement.

In the pilot system, existing MT technologies are tested in different scenarios. Currently delivered are the best evaluated ones. Not all scenarios or usage models are accepted by users. Users can easily get frustrated and give up using the services.

#### **5 Conclusion**

The multilingual information service system is important to Beijing Olympic Games. MT is going to play an important role among other technologies. The pilot systems have revealed that under carefully designed usage models, MT is functioning well and positively accepted.

The system is not built only for the Olympics, it will also serve the general public and foreign visitors before and after Olympics. The technology solutions and know-how accumulated will be used in similar systems and events.

When building such a complicated information system, especially where multilingual MT technologies need to be improved greatly and in a short time period, international collaborations are indispensable and are always encouraged.