International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.5 (2016), pp. 11-20
http://dx.doi.org/10.14257/ijmue.2016.11.5.02

Design of a Subtitle System for Internet TV Systems

Jaegeol Yim

[1]*Department of Computer Engineering, Dongguk University at GyeongJu, 38066, Korea*
yim@dongguk.ac.kr

Abstract

Globalization was one of the most popular words in Korea about 20 years ago. Koreans believed that a nation cannot join the ranks of developed countries unless that nation is globalized. In order for Internet TV systems to be globalized and to spread throughout the world, broadcasters must provide a subtitling service in foreign languages. Otherwise, their programs cannot be understood by foreigners. There are authoring tools and automatic translators that can generate subtitles. This paper reviews techniques needed in subtitle system development, and introduces the design of a subtitle system for Internet TV. One of the distinguishing features of the design is that a machine translator is integrated into the system. A crawler is proposed as part of the translator to collect cultural and tourism proper nouns.

***Keywords**: Internet TV, Subtitle, Natural Language Translation, Application Programming Interface*

1. Introduction

Subtitles are transcripts of the dialog in television programs and are very useful. They are necessary for audience members who are deaf, and subtitles in a foreign language can help students to learn the language. Besides, globalization cannot be achieved unless the language barrier is removed, and subtitles can be used to mitigate the barrier.

Globalization is the key word for all developing countries. Only after becoming globalized can a developing country join the ranks of developed countries. In order to globalize Internet TV, broadcasters have to provide subtitles in English or Chinese [1].

This paper surveys existing subtitle systems and designs a new subtitle system for Internet TV. The system uses translation cloud services. There are many large companies providing translation services. Application programming interfaces (APIs) provided by these machine translation service providers are discussed.

2. Related Work

Internet TV, natural language processing, machine translation, and subtitles are the topics related to this paper. Given a manuscript, an article, a set of sentences or phrases, or a speech, recognizing the author's or the speaker's personal information and attitude is very important. With this personal information, politicians can make better campaign plans, and company decision makers can establish better marketing strategies. Sentiment analysis is the field of research where techniques to recognize an author's attitude and personal information are studied, and it has been a hot research topic in natural language processing (NLP). Bhattacharyya *et al*. [2] noticed that the main subjects that researchers of sentiment analysis pursue include subjectivity detection, polarity detection and degree of polarity identification. They have summarized techniques used to develop a system that extracts subjective sentences from given documents. A translation support system consisting of a morphological analyzer, a lexical dictionary, a corpus, a rule base, pattern-

ISSN: 1975-0080 IJMUE

directed parsing, speech recognition, and text-to-speech synthesis, among other components, was discussed [1].

Chandak and Dharaskar [3] described a home appliance that records a user's voice commands ("open" and "close," for example). It understands the user's commands and performs the instructions. The system is also equipped with cameras that capture images, sending them to the server. The server analyzes the images and sounds an alarm if an abnormal situation is detected [1].

If I post the sentence, "They took pictures of me and my classmates around the Wonhyo Building on Oct. 15 for the photo album of the graduates," someone might figure out which university I am attending and how old I am. Kataoka *et al*. [4] developed Disclosure Control of Natural Language (DCNL), which automatically changes a sentence that reveals private information into another sentence that conceals that information. For example, DCNL might change the above sentence into "My friends and I took pictures of ourselves around the main building of the university" [1].

As someone speaks in English, the Skype Translator app displays subtitles in both English and German. This implies that the app recognizes spoken English, translates English into German, and provides speech synthesis. The Microsoft translator applies phrasal statistical machine translation (phrasal SMT) to translate one natural language into another. For example, it has texts that were translated from English into German. When an English sentence is given, it chops up the sentence into a set of phrases. It searches for each phrase in the translated texts, and returns the German version. Then it composes all returned phrases to come up with a German sentence. Recently, the Microsoft translator used syntactically informed phrasal statistical machine translation (or syntactic SMT). Syntactic SMT maps each English word to a German word. This strategy seems like a primitive approach. However, syntactic SMT is much better than phrasal SMT [5].

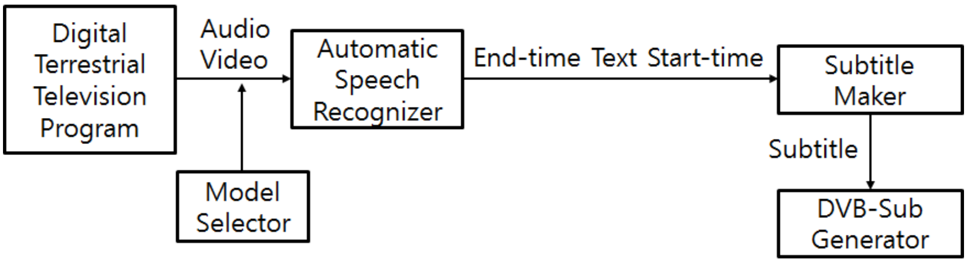


Figure 1. Software Modules for Subtitle Generation [6]

Castro *et al*. [6] introduced a real-time subtitle generator where input was a terrestrial television program and output was an MPEG transport stream for an IPTV channel. A digital terrestrial television program can be classified into news, sports, a news magazine, and so on. The model selector recognizes the type of TV program so the appropriate automatic speech recognizer (ASR) can be selected. The selected ASR transcribes the audio into text. After augmenting the resulting text with the start time and end time of the audio, the ASR transmits the string to the subtitle maker. The subtitle maker converts the received string into subtitles and transmits the subtitles to the digital video broadcasting (DVB)-sub generator as shown in Figure 1 [6].

The DVB-sub generator creates the packets containing the subtitles, DVB-Subtitles, as shown in Figure 2. While the audio is transcribed and processed to become DVB-Subtitles, the audio and video data are buffered and wait until DVB-Subtitles is generated. Finally, the muxer combines the packetized elementary stream (PES) of audio and video, DVB-Subtitles, and the program map table (PMT) in order to create an MPEG transport stream, as shown in Figure 2 [6].

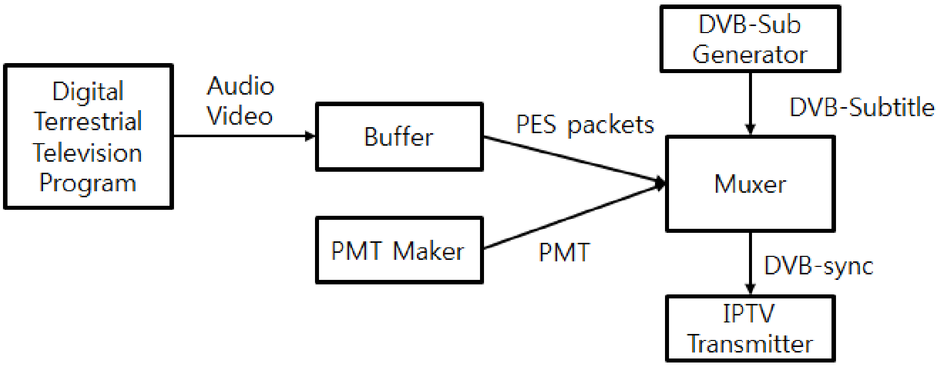


Figure 2. Data Flow for Subtitle Synchronization [6]

Orero *et al*. [7] described the Hybrid Broadcast Broadband for All (HBB4ALL) system, in which broadcast and broadband services converged. For subtitling, this system provides to users and automatic speech recognizers authoring tools developed by partner systems. The subtitling component of the system is compatible with the following standards and specifications: Tech 3350, Tech 3360, Tech 3264-1991, Timed Text Markup Language 1 (TTML1), and New for Protocol for Teletext Subtitles. One of the main distinguishing features of this subtitling service is that it allows users to change the font, the color, the location of subtitles, the duration of the subtitles, and the temporal presentation style [7].

This system also provides audio description in the form of spoken commentary. The audio description contains not only semantic information but also affective prosody. This system uses text-to-speech (TTS) systems to generate the audio description in purely informative genres. HBB4ALL also provides audio introductions. An audio introduction provides information about the TV program and includes the running time, characters, locations, costumes, and production credits. They expect that audio introduction will be used by many people who are visually impaired [7].

A focused crawler is an Internet bot that collects web pages only if they are closely related to the specified theme. Xin and Qiuli [8] introduced a focused crawler for business intelligence acquisition. The process performed by the crawler is described in Figure 3. As shown in the figure, it saves only related Web pages.

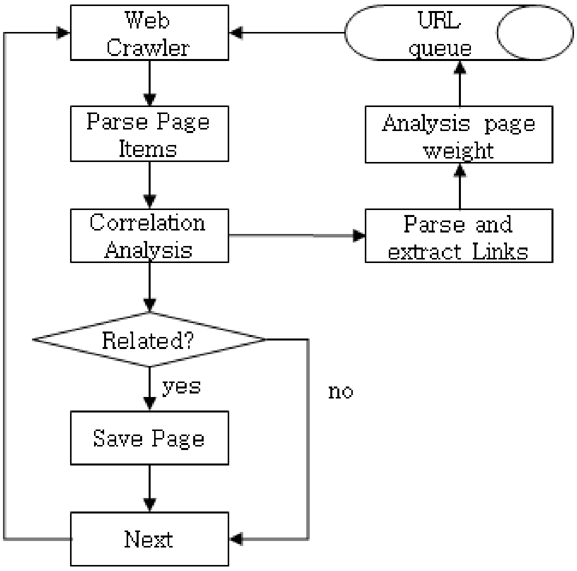


Figure 3. The Process Focused Crawlers

3. Design of an IPTV System

One of the key elements of our subtitle system is the translator. The translator converts Korean to English and Chinese. Many translators have been tested, and there is no perfect Korean translator. One of the shortcomings of existing translators is that they do not properly translate proper nouns. In order to mitigate this shortcoming, preprocessing the input is suggested. The preprocessor identifies proper nouns in the input and properly translates them, referring to the dictionary of proper nouns shown in Figure 4. Given a text, the translator invokes the preprocessor, passing the text as an argument. The preprocessor returns the text after replacing proper nouns with the designated foreign language words. Using this text as an argument, the translator invokes the application program interface (API) of the translation service. The text written in the foreign language generated by the translation service is passed to the IPTV subtitle service.

The dictionary of proper nouns provides a user interface through which users can conveniently add, update, and retrieve proper nouns, as shown in Figure 5. Users can add more words by clicking the + button. Before adding words, users can search for them in the dictionary by clicking the magnifying glass button. The managers of the system can retrieve all user input and edit it as needed. Only managers are allowed to add new entries to the dictionary.

Another source of collecting proper nouns is the Web crawler. A Web crawler is a program that visits web pages designated by the manager and collects information, copying the whole page, as shown in Figure 6. Then, the natural language processing service analyzes the web page and collects proper nouns. After copying the designated page, the crawler moves to another page following the links defined in the visited page.

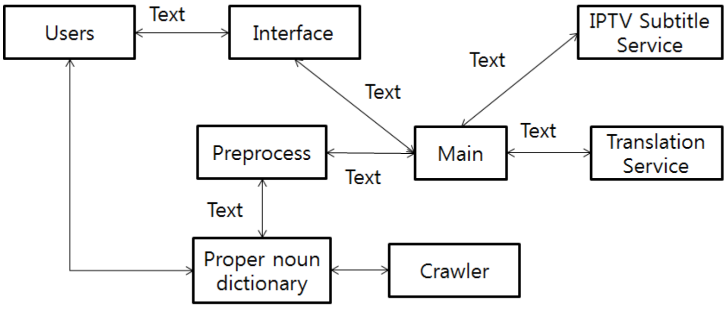
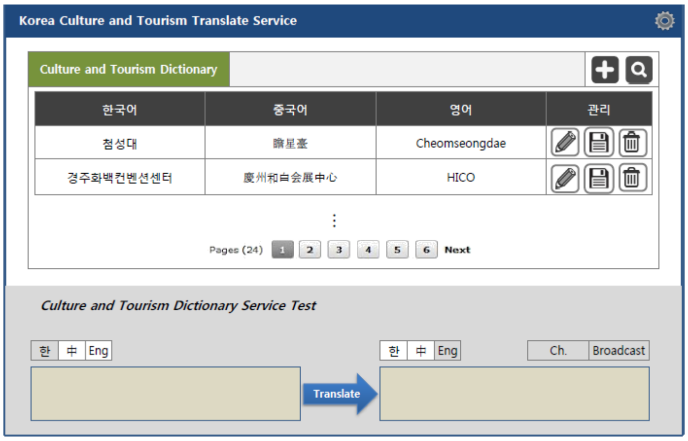


Figure 4. Data Flow for Preprocessing

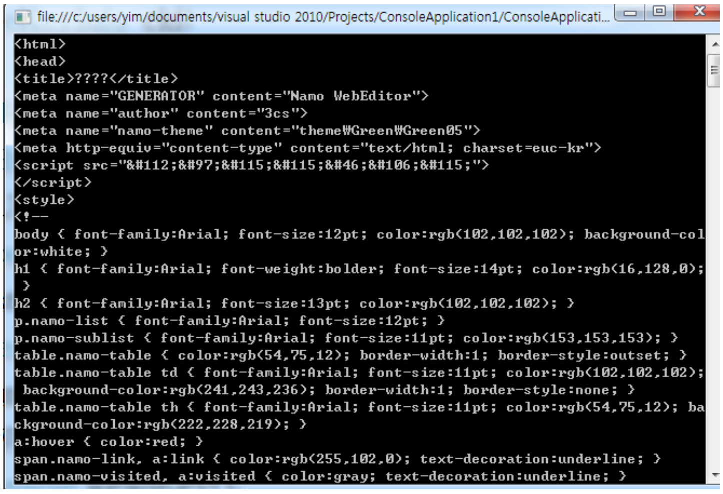

Korea Culture and Tourism Translate Service
Culture and Tourism Dictionary

한국어	중국어	영어	관리
첨성대	瞻星臺	Cheomseongdae	
경주화백컨벤션센터	慶州和白会展中心	HICO	

⋮

Pages (24) 1 2 3 4 5 6 Next
Culture and Tourism Dictionary Service Test
한 中 Eng
한 中 Eng
Ch.
Broadcast
Translate

Figure 5. An Example User Interface to Collect Proper Nouns


file:///c:/users/yim/documents/visual studio 2010/Projects/ConsoleApplication1/ConsoleApplicati...

```
<html>
<head>
<title>????</title>
<meta name="GENERATOR" content="Namo WebEditor">
<meta name="author" content="3cs">
<meta name="namo-theme" content="theme₩Green₩Green05">
<meta http-equiv="content-type" content="text/html; charset=euc-kr">
<script src="&#112;&#97;&#115;&#115;&#46;&#106;&#115;">
</script>
<style>
<!--
body { font-family:Arial; font-size:12pt; color:rgb(102,102,102); background-col
or:white; }
h1 { font-family:Arial; font-weight:bolder; font-size:14pt; color:rgb(16,128,0);
 }
h2 { font-family:Arial; font-size:13pt; color:rgb(102,102,102); }
p.namo-list { font-family:Arial; font-size:12pt; }
p.namo-sublist { font-family:Arial; font-size:11pt; color:rgb(153,153,153); }
table.namo-table { color:rgb(54,75,12); border-width:1; border-style:outset; }
table.namo-table td { font-family:Arial; font-size:11pt; color:rgb(102,102,102);
 background-color:rgb(241,243,236); border-width:1; border-style:none; }
table.namo-table th { font-family:Arial; font-size:11pt; color:rgb(54,75,12); ba
ckground-color:rgb(222,228,219); }
a:hover { color:red; }
span.namo-link, a:link { color:rgb(255,102,0); text-decoration:underline; }
span.namo-visited, a:visited { color:gray; text-decoration:underline; }
```

Figure 6. An Example Web Page Copied by a Crawler

Considering the real-time subtitle generator and the translator, the proposed subtitle system was designed for live programs, as shown in Figure 7. The system is essentially a combination of the real-time subtitle generator and the translator shown in Figures 1, 2, and 4. The automatic speech recognizer generates text augmented by the start time and the end time for synchronization. The translator provides APIs that take this text as input and translate it into the target language.

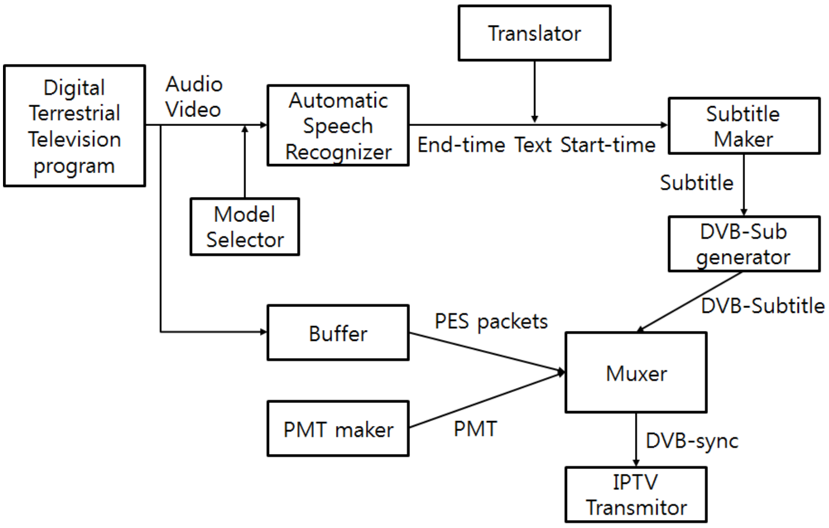
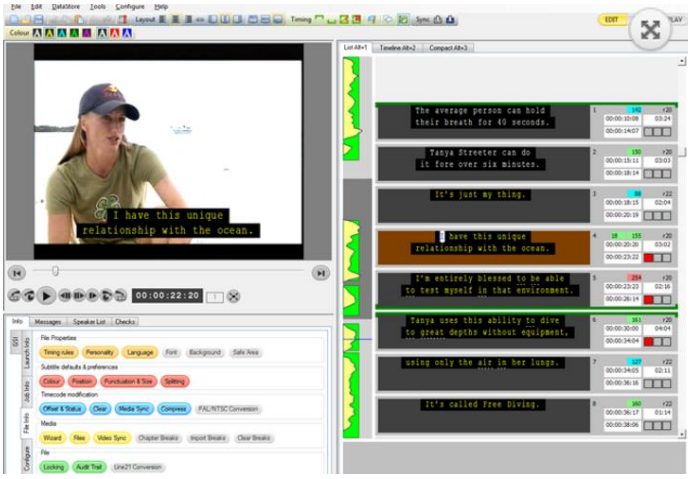


Figure 7. Design of the Proposed Subtitle Generator for Live Programs



I have this unique
relationship with the ocean.

00:00:22:20

1 142 r20 00:00:10:08 03:24 00:00:14:07
The average person can hold
their breath for 40 seconds.

2 150 r20 00:00:15:11 03:03 00:00:18:14
Tanya Streeter can do
it fore over six minutes.

3 88 r22 00:00:18:15 02:04 00:00:20:19
It's just my thing.

4 18 155 r20 00:00:20:20 03:02 00:00:23:22
I have this unique
relationship with the ocean.

5 254 r20 00:00:23:23 02:16 00:00:26:14
I'm entirely blessed to be able
to test myself in that environment.

6 161 r20 00:00:30:00 04:04 00:00:34:04
Tanya uses this ability to dive
to great depths without equipment.

7 127 r22 00:00:34:05 02:11 00:00:36:16
using only the air in her lungs.

8 160 r22 00:00:36:17 01:14 00:00:38:06
It's called Free Diving.

Figure 8. User Interface of the Wincaps Subtitling Software [9]

Subtitles can be created, and there are many authoring tools that can conveniently generate subtitles. For example, the WinCaps subtitling software provides a stylish and intuitive user interface, as shown in Figure 8 [9]. Users can conveniently specify subtitles, the start time and end time of the display, the background color, the font, and so on, through the interface. The authoring tools should support translation so that subtitles can be written in many languages. Plans include enhancing the proposed translator so it takes subtitles from the authoring tool and returns translations, as shown in Figure 9.

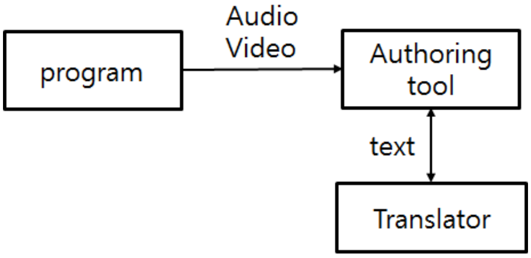


Figure 9. The Translator Interacts with the Authoring Tool

4. Implementation Considerations

Using the Microsoft translator APIs, software developers can create web and client applications that translate one language to another. The Microsoft translator provides a collaborative translations framework (CTF) and the Microsoft translator hub (MTH). Using CTF, users can correct the translation result generated by the Microsoft translator. As users make corrections, the accuracy of the translator improves. Using MTH, users can build their own dictionaries. Web application developers can easily enhance their web pages with translation services by using the translator web widgets provided by Microsoft developer networks. Application developers can use the translation API through AJAX, HTTP and SOAP [1].

Before using the Microsoft translation API, users have to visit https://datamarket.azure.com/home and subscribe to the Microsoft Azure Marketplace with a Microsoft account. Then, it is necessary to visit www.aka.ms/TranslatorADM in order to have a translation API plan. There is an API plan that allows users to freely translate up to 2 million characters per day. When subscribed to an API plan, users have to memorize "clientID" and "client secret" to enter when accessing the translation APIs [1].

Whenever a translation API is invoked, it is necessary to pass an access token. This access token is also needed for user account management. In order to obtain an access token, users subscribe to a translation API, register the application being developed, and send an HTTP POST request to the token service. Figure 10 shows part of the source code introduced in https://msdn.microsoft.com/en-us/library/hh454950.aspx, which obtains an access token and uses the Detect API function [10].

The process to execute this program is as follows. Visit http://www.microsoft.com/en-us/download/details.aspx?id=30653 to download and install the .NET 4.5 full setup. Then, execute Visual Studio. In the Solution Explorer window of Visual Studio, click References with the right mouse button, select Add Reference…, and add System.Runtime.Serialization and System.ServiceModel. Then add System.Web, which requires.NET Framework 4 or above. The process of setting a target framework can be summarized as follows: Project -> project name Properties -> Application -> Target framework [1].

The “clientID” and “client secret” parameters were obtained earlier. If they are wrong, then the system generates error code 400. This program creates an AdmAuthentication object with “clientID” and “client secret.” Then, it generates a request to http://api.microsofttranslator.com/v2/Http.svc/Detect, adds an access_token to the header of this request, and finally requests the service [1].

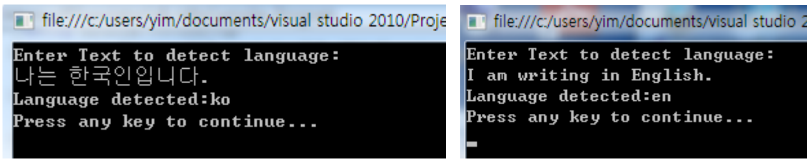
Execution results of the program shown in Figure 10 are shown in Figure 11. If entering Korean characters, then the program responds with “ko”; entering English characters prompts a response with “en” instead.

In order to translate English into Korean, change the URI in Figure 1 to "http://api.microsofttranslator.com/v2/Http.svc/Translate?text=" + System.Web.HttpUtility.UrlEncode(text) + "&from=" + “ en” + "&to=" + "ko"; After entering “I am an American” in the text variable, executing the program obtains the result shown in Figure 12.

```
    using System;
    ...
    using System.Threading;
    AdmAuthentication admAuth = new AdmAuthentication("clientID", "client secret");
    ...
    admToken = admAuth.GetAccessToken();
    headerValue = "Bearer " + admToken.access_token;
    DetectMethod(headerValue);
    ...
    Private static void DetectMethod(string authToken)
    {
    Console.WriteLine("Enter Text to detect language:");
    string textToDetect = Console.ReadLine();
    string   uri   =   "http://api.microsofttranslator.com/v2/Http.svc/Detect?text="   +
textToDetect;
    HttpWebRequest httpWebRequest = (HttpWebRequest)WebRequest.Create(uri);
    httpWebRequest.Headers.Add("Authorization", headerValue);
```

```
...
```

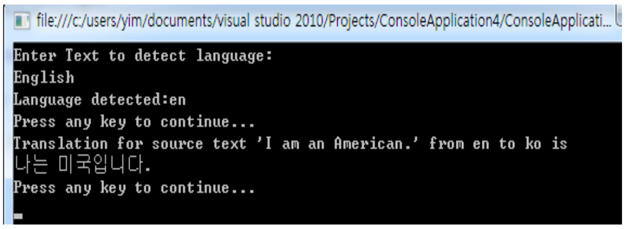
Figure 10. Part of the Program that Detects the Input Language


file:///c:/users/yim/documents/visual studio 2010/Proje
file:///c:/users/yim/documents/visual studio 2

```
Enter Text to detect language:
나는 한국인입니다.
Language detected:ko
Press any key to continue...
```

```
Enter Text to detect language:
I am writing in English.
Language detected:en
Press any key to continue...
```

Figure 11. Test Results of the Program in Figure 10


file:///c:/users/yim/documents/visual studio 2010/Projects/ConsoleApplication4/ConsoleApplicati...

```
Enter Text to detect language:
English
Language detected:en
Press any key to continue...
Translation for source text 'I am an American.' from en to ko is
나는 미국입니다.
Press any key to continue...
```

Figure 12. A Translation Result

Using the HttpWebRequest class provided by .NET, users can access any web page. For example, the program shown in Figure 13 creates a request for "http://wwwcs.dongguk.ac.kr/~yim/" and obtains a copy of the page by invoking the GetResponse() method. Then, it turns the page into a stream with GetResponseStream() and gets the text version of the stream with the ReadToEnd() method [11].

```
...
ConsoleKeyInfo cki;
HttpWebRequest request = (...)HttpWebRequest.Create("http://wwwcs.dongguk.ac.kr/~yim/");
request.UserAgent = "A .NET Web Crawler"
WebResponse response = request.GetResponse();
Stream stream = response.GetResponseStream();
StreamReader reader = new StreamReader(stream);
string htmlText = reader.ReadToEnd();
Console.Write(htmlText);
cki = Console.ReadKey();
...
```

Figure 13. Part of the Web Crawler Source Code

5. Conclusions

Subtitles are necessary for globalization. Subtitles written in foreign languages should be provided for TV programs. This paper introduced the design of a subtitle system. The system is integrated with a machine translator. The proposed translator is equipped with a dictionary of proper nouns. A user interface for the dictionary was discussed. The translator interacts with existing authoring tools. That is, the translator should be able to take subtitles in the mother tongue, translate them into target languages, and return the resulting translations to the authoring tool.

Acknowledgements

This research was funded by the Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2011-0006942), by the 'Development of Global Culture and Tourism IPTV Broadcasting Station' Project through the Industrial Infrastructure Program for Fundamental Technologies funded by the Ministry of Knowledge Economy (10037393), and by the Dongguk University Research Fund of 2016.

References

[1] J. Yim, "Design of a Subtitle Generator", ASTL, **(2015)**, vol. 117, pp.73-76.

[2] D. Bhattacharyya, S. Biswas and T. Kim, "A Review on Natural Language Processing in Opinion Mining", IJSH, **(2010)**, vol. 4, no. 2, pp. 31-38.

[3] M. Chandak and R. Dharaskar, "Natural Language Processing based Context Sensitive, Content Specific Architecture & its Speech based Implementation for Smart Home Applications", IJSH, **(2010)**, vol. 4, no.2, pp. 1-10.

[4] H. Kataoka, N. Watanabe, K. Mizutani and H. Yoshiura, "DCNL: Disclosure Control of Natural Language Information to Enable Secure and Enjoyable E-Communications", IJUNESST, **(2010)**, vol. 3, no. 1, pp. 1-10.

[5] T. Chong, "How social media teaches skype to speak [News]", IEEE Spectrum, vol. 51, no. 9, **(2014)**, pp. 15-16.

[6] M. de Castro, D. Carrero, L. Puente and B. Ruiz, "Real-time subtitle synchronization in live television programs", Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), June 08-10, Erlangen Germany, **(2011)**, pp. 1-6.

[7] P. Orero, C. A. Martin and M. Zorrilla, "HBB4ALL: Deployment of HbbTV services for all", Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), June 17-19; HetPand Ghent, Belgium, **(2015)**, pp. 1-4.

[8] P. Xin and Q. Qiuli, "Focused Crawler Research for Business Intelligence Acquisition", IJHIT, vol. 6, no. 6, **(2013)**, pp. 187-194.

[9] http://www.screensystems.tv/products/wincaps-subtitling-software/

[10] https://msdn.microsoft.com/en-us/library/hh454950.aspx

[11] http://ericsowell.com/blog/2007/8/14/how-to-write-a-web-crawler-in-csharp

Author

Jaegeol Yim, received an MSc and PhD in Computer Science from the University of Illinois at Chicago, in 1987 and 1990, respectively. He is a Professor in the Department of Computer Engineering at Dongguk University at Gyeongju in Korea. His current research interests include Petri net theory and its applications, location-based services, AI systems, and multimedia systems. He has published more than 50 journal papers, 100 conference papers (mostly written in the Korean language), and several undergraduate textbooks.