

Multilingual Web with E-speranto

Tomazič, Sašo

Abstract— *This paper presents the idea of E-speranto, a Hyper Text Description Language (HTDL) which will act as an extension of Hyper Text Markup Language (HTML) for multilingual documents. Documents written in E-speranto will be interpreted in Web browsers in a chosen natural language, and thus anyone will be able to read these documents in his/her own language. This will eventually bring an end to the "language divide".*

Index Terms— *Esperanto, language, multilingual, translation, Web, interpreter.*

1. INTRODUCTION

IN the mid-1990s the term "digital divide" came into regular use to refer to the gap between those with regular access to digital and information technology, and those without this access. Now, seventeen years later, the digital divide is closing. Never before there have been so many people with access to electronic communication technologies and the Internet. However, another divide is becoming more and more apparent. We can call it the "language divide". It divides those who understand English and those who do not.

Virtually any kind of information can be found on the World Wide Web (WWW), available to any one who can understand it. In most cases, Web pages are in English, or at least their English translations are available. So, understanding English is a necessary prerequisite for finding desired information on the Web; however, it is not sufficient. It can often happen that the information we are searching for only exists in a regional language we do not understand.

The digital divide was relatively easy to overcome. All that was necessary was to develop new technologies, cheap enough to be affordable for everyone. On the other hand, it is much more difficult to teach everyone English, and it is clearly impossible to teach everyone every language in the world.

One avenue to the goal of overcoming language barriers and bringing the language divide to an end is the use of automatic translation tools. There are already many of them available on the Web. Some of them can even be used free of charge for on-line translation of

visited pages or for translation of text inserted in a text input box. The main drawback of this approach is that these tools only work with pairs of languages, and only major world languages are supported. Google, for example, offers translations between English and Arabic, Chinese, French, Italian, Korean, Japanese, Russian, Spanish and Portuguese. The only pair that does not include English is from German to French and vice versa.

If you do not happen to understand some of the above languages, or even if you do and come across a page written in another language, Slovenian for example, you are doomed. It is not very likely that automatic translation between pairs of marginal languages, e.g. Slovenian to/from Vietnamese, will be available in the foreseeable future. Moreover, in order to have automatic translation between all pairs of 6700 existing languages, a total of 44,883,300 compilers would be needed (two for each pair).

A much better solution to this problem can be found in the world of computer communications. Each computer understands its own machine language; however, there is no need to understand machine languages of other computers in order to communicate with them. All that is needed is a common language – a standardized communication protocol – which is then locally interpreted on each computer in a language understandable to the computer processor.

With regard to the Web, Hyper Text Markup Language (HTML), together with its underlying communication protocols, is an excellent example of such a language. There is no need for documents on the Web to be in a format specific to a certain computer or operating system as long as a suitable HTML interpreter (Web browser) is installed on the computer. The interpreter reads the HTML documents and interprets them in a language understandable to the computer-specific graphics card in order to be shown on that computer's specific screen.

The introduction of HTML was a success story – the World Wide Web was born. Many extensions to HTML have been made since then to enable a richer design of documents and to make the design process easier. However, HTML is still a language that only allows interpretation of the document format. The text itself remains in its original language.

It is now time to take the next logical step: to design a common Hyper Text Description Language (HTDL) which will allow interpretation

in different natural languages. Documents written in HTML/HTDL will not only be correctly shown on the computer screen but will also be interpreted in a language understandable to the user browsing them.

2. E-SPERANTO

To avoid tedious use of the acronym HTDL for Hyper Text Description Language, I have named it E-speranto (pronounced i:spærae'ntou). The name is derived from Esperanto (pronounced espærae'ntou) [1]. Esperanto is a constructed international auxiliary language proposed by L. L. Zamenhof in 1887 as a universal second language that would enable people of different nations to communicate and understand each other. E-speranto can be considered an electronic version of Esperanto.

Although the concepts of E-speranto and Esperanto share a common goal, there is one substantial difference between them. To understand Esperanto one must learn it, whereas it would be sufficient to have an interpreter installed on a computer to understand E-speranto. This difference can make E-speranto succeed where Esperanto failed.

As already mentioned, Esperanto was intended as a universal second language. To this end, everyone should learn it. However, learning a new language is a difficult and time-consuming task. One usually decides to learn a new language only when she/he feels that it would be very useful in her/his business or private life. As only a relatively small number of enthusiasts currently speak Esperanto, one cannot expect to be able to conduct business in Esperanto nor to be understood when ordering food or buying clothes. The critical mass of Esperanto speakers was not reached.

On the other hand, interpreters for E-speranto could be integrated in Web browsers so that anyone would be able to see documents written in E-speranto in her/his own language without even being aware that they are written in E-speranto. Only one interpreter needs to be developed for each language. For small nations and/or small ethnic groups it would be much more feasible to develop a single E-speranto interpreter than to develop compilers for all languages of interest.

3. DESIRED PROPERTIES

E-speranto does not exist. It has yet to be designed. To be successful and to best serve its purpose, certain properties are desirable and should be considered when designing E-speranto. These properties are the following:

1. A meaning that can be expressed in a natural language can also be expressed in E-speranto. This is the basic requirement for any synthetic language for communication among people and, as such, is also a property of

Esperanto.

2. As E-speranto is intended to be interpreted by computers, its syntax should be computer friendly. This implies that all words in E-speranto have only one form and that grammar is expressed explicitly in the form of tags.

3. E-speranto documents will be mostly designed by humans. Tools for editing E-speranto documents and E-speranto interpreters will also be developed by humans. Thus, E-speranto documents should be readable by humans. For this purpose, eXtended Markup Language (XML)-based syntax seems a good choice.

4. To simplify the development of interpreters, the grammar of E-speranto must have no exceptions. For that purpose it can be based on the grammar of Esperanto, only that it must be expressed explicitly (by tags), not implicitly (by prefixes, suffixes, etc.) as in Esperanto.

5. E-speranto vocabulary should contain unique concepts, i.e. words with a single meaning and no synonyms. In natural languages, some words represent many different concepts and some concepts can be represented with many different words. Both must be avoided in E-speranto. E-speranto vocabulary can also be based on Esperanto, but probably English is a better choice, because more developers (computer programmers) understand English than Esperanto.

6. The meaning and the style of the language must be separated. Style can be expressed in some kind of translation hint, which would tell the interpreter which style to use for interpreting the text of the document (e.g. dialect, slang, poetical, etc.). If all translation hints are ignored, the basic meaning of the text must remain. This is similar to the separation of document structure in HTML and style in CSS.

4. CONCLUSION

According to the above-presented idea of HTDL we embarked on development of E-speranto at the Faculty of Electrical Engineering, University of Ljubljana. The home page of E-speranto [3] is under construction and will be available soon. We would initially like to design E-speranto syntax and grammar in such a way that a human would be able to translate documents written in E-speranto to his/her language. If a human cannot do it, a computer cannot do it. In parallel we began development of tools for writing documents in E-speranto and also began researching algorithms for interpreting E-speranto in three Slavic languages: Slovenian, Serbian and Russian. The last and most difficult phase will be the development of translators from natural languages to E-speranto. When and if such translators are developed for different natural languages, any document would be easily translated from/to any of those languages using E-speranto as an intermediate

language.

I hope that this or other similar efforts (e.g. [4]) will bear fruit and that some form of HTDL will be standardized to break down language barriers and close the language divide, thus bringing different nations closer together.

ACKNOWLEDGMENT

It was largely owing to my colleague and friend Franci Demšar that I came to the idea of E-speranto. He, not being from the field, was convinced that automatic translation from some kind of simplified language into different natural

languages should be an easy task in the age of computers. This brought the idea of using Esperanto as a basis of such a language.

REFERENCES

- [1] F. Amerio, G. Bonvecchiato, G.C. Fighiera (by), "Esperanto: Data and Facts. 2nd edition," *FEI-Milan 2002*.
- [2] S. Omerovič, G. Jakus, T. Filomonova, S. Tomažič. "Multilingual documents in e-speranto," *Elektroteh. vestn.*, Vol. V4, No. 4, pp. 151-157, 2007.
- [3] <http://www.e-speranto.com>.
- [4] <http://www.undl.org>.