Information Management Resource Kit

Module on Building Electronic Communities and Networks

UNIT 4. DESIGNING AN ONLINE COMMUNITY

LESSON 6. MULTILINGUAL COMPUTING

NOTE

Please note that this PDF version does not have the interactive features offered through the IMARK courseware such as exercises with feedback, pop-ups, animations etc.

We recommend that you take the lesson using the interactive courseware environment, and use the PDF version for printing the lesson and to use as a reference after you have completed the course.

Objectives

At the end of this lesson, you will be able to:

- define the range of issues relating to multilingual computing and linguistic diversity;
- describe the potential role of open source software tools in addressing language issues;
- identify the steps to follow in order to plan and implement a language-related strategy for your online community;
- understand the main strengths and limitations of online/machine translation tools and be able to use an online translation tool effectively.

Introduction

The term **multilingual computing** covers a wide range of issues facing online communities and networks, from non-Latin alphabets (such as Chinese or Arabic) to developing Web sites for multilingual audiences.

There are two main types of issue:

- **technical issues**, concerning how user-friendly the software and access to computing is for people from different language groups; and
- **communication issues**, concerning how well these users can understand the content of communication, and how well they are understood when they make a contribution.

Introduction

Computing developed in a predominantly English-speaking environment that uses Latin characters without (or with few) accents, and computing technologies are still best adapted to English.

The issues facing your online community will depend on the particular languages used by community members.

The less used a language is, and the less common the fonts, the more challenges speakers of those languages will face.

Introduction

If your online community is **monolingual**, and from a dominant language group (such as English or French) you will have few if any language-specific problems.

You will find a range of high-quality software available in your group's language, and you will not face challenges of cross-language communication.

On the contrary, if your community is **multilingual**, and/or it communicates in one or more languages which use non-Latin character sets (alphabets), you are likely to face a range of challenges relating to multilingual computing and communication.

For example,

it may be difficult to find e-mail clients or word processing applications in the languages used by the community, or members may have to agree on a primary language of communication for the group which is not the first language of most participants.

See interactive lesson for some issues relating to multilingual computing

Technical issues: software interface language

The two main technical issues relating to multilingual computing are:

1. **software interface language**, and
2. **fonts and encoding**.

The **software interface language** is the language used for the menus, commands, help functions etc.

The vast majority of software applications produced worldwide have an **English-language interface**. Although major software packages are increasingly available in **other language versions**, these are usually only a few dominant languages such as **French**, **German**, and **Spanish**.

People who do not understand these languages can either learn to operate the software "by heart", using commands that they don't understand, or more likely, will not be able to use it at all.

Technical issues: software interface language

Developers of both commercial and open source software are increasingly focusing on making software available in a multitude of languages in order to meet the needs of users who speak languages that belong to a small language group.

With regard to **open source** software, availability and usefulness varies greatly from language to language. You can try to network with persons of similar concerns on the Internet, to find ever changing information about such software or tools for small compromises while using generally available software.

A growing number of open source applications, such as OpenOffice.org or Mozilla Thunderbird and Mozilla Firefox, also include spell checkers for a range of languages.

The Khmer-language e-mail open source software

Technical issues: software interface language

Open Office

Open Office is an open source office suite that provides functions similar to Microsoft's **Office** application suite (word processor, spreadsheet, presentation tool).

Like other open source applications, Open Office is available free of charge, currently in approximately 40 language versions, including such languages as African Zulu, Xhosa, Asian languages Tamil, Lao and many others.

A full list of language versions is available from http://download.OpenOffice.org/

Mozilla

Examples of an e-mail client and a web browser available in multiple languages, are **Mozilla Thunderbird** and **Mozilla Firefox** (localized into over 70 languages). A full list of language versions is available from http://www.mozilla.org/projects/l10n/mlp_otherproj.html#ffox_contrib.

Technical issues: software interface language

Khmer-language open source software

Until now, a native Cambodian has needed to be able to read in a foreign language to be able to send e-mail in Cambodia. Software was not available in Khmer.

Now things have changed, the *Khmer Software Initiative* (http://www.khmeros.info/drupal), set up to bring together previously isolated developers, is starting out by providing **Khmer translations** of well-known free applications such as the powerful **e-mail application**, **Mozilla Thunderbird**.

Technical issues: software interface language

What multilingual computing challenges is a user from a small language group likely to face?

- ☐ Difficulties in finding software with an interface/operating commands in their language.
- ☐ Challenges related to time-zone differences when communicating with people on the other side of the globe.
- ☐ Need to use foreign language for communicating outside his/her community.
- ☐ Font-related challenges when using a Latin character set for communication.
- ☐ Font-related related challenges when using non-Latin alphabet for communication.

Please select the answers of your choice (2 or more) and press Check Answer.

Technical issues: fonts and encoding

If you are working in a multilingual group you need software which:

- can process characters in **your language**; and
- can display characters from **other languages**.

Even if your network's communication is primarily in one language, it may be necessary to display characters from other languages.

For example, even if English is the primary medium of communication in your group, your communications might contain French terms or names written in their original form (e.g. "Côtes du Rhône" or "Agence Française de Développement").

The larger the number of characters used in a language which fall outside the set used in dominant languages such as English and Spanish, the more problems you will face.

The linguistic environment for computer work

Computers for communication in writing were originally developed in the USA, in an environment with an extremely simple writing system: the Latin alphabet with no accented letters, and with no semantic differences between lower case and upper case letters. This simple set of letters is called the "**ASCII standard**," or more precisely, the "American Standard Code for Information Interchange," a standard code that was proposed by the American National Standards Institute in 1963, and finalized in 1968.

All languages that require either modifications or additions to the English-language alphabet, or even different letters to be written (e.g. the Spanish ñ) require different sets of standards. One standard which has been developed to address this problem is **Unicode**.

The Unicode Standard is a character coding system "designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world."

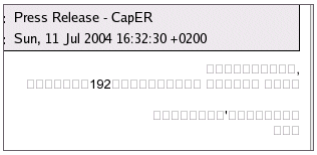
For more information about Unicode see the Unicode site: http://www.unicode.org/

Technical issues: fonts and encoding

E-mail communication using non-Latin scripts

Two main types of problem face e-mail users:

- the *sender* of the message needs to have an e-mail program which enables them to write messages using all the characters they need to; and
- the *receiver* of a message needs to have an e-mail program which *displays* these characters correctly. If it does not, the recipient may receive a message similar to the one below…



Technical issues: fonts and encoding

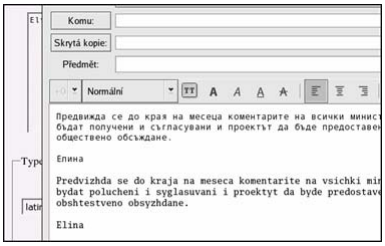
What strategy could you use to ensure that the recipient of your e-mail will be able to read your message, if you use non-Latin characters for writing?

- ○ Send the message as plain text (not HTML message).
- ○ Send with the text of the message included as an e-mail attachment written in MS Word or other word processor.
- ○ Send your e-mail encrypted.

Please click on the answer of your choice.

Technical issues: fonts and encoding

E-mail communication using non-Latin scripts



e-mail message in two formats

You can minimize the risk that your information won't be delivered in the correct format by:

- **writing the message in a word processing application that supports the language** (for example, OpenOffice.org Writer or MS Word) **and attaching it to the e-mail message**; and
- **sending messages in two formats**, one using the original alphabet and other using Latin characters without accents

(see more information on next slides).

Typing in non-Latin characters

Technical issues: fonts and encoding

Writing a message in a word processing application that supports the language and attaching it to the e-mail message

While attachments to e-mail messages (particularly when sent to mailing lists) can be problematic (for example, due to the potential for transmitting viruses, or the size of attachments), this **may be the only way to work around the problem of unsupported character sets**.

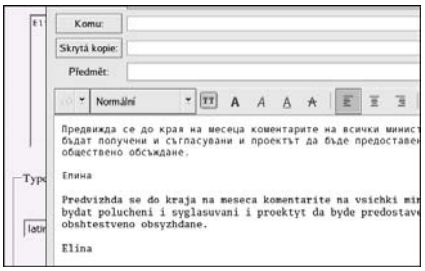
This strategy is often used when distributing documents such as press releases or e-newsletters, through distribution lists, whose recipients are running diverse applications, ranging from web based mail to open source and commercial e-mail clients.

Technical issues: fonts and encoding

Sending messages in two formats: one using the original alphabet and other using Latin characters without accents

This approach is sometimes used for messages in Cyrillic or in Greek where letters can be transcribed into Latin and vice versa. Language converter software can be used to convert between Latin and Cyrillic or Greek. Similar converters exist to convert documents in Latin characters with special accents into plain text (text without accents).

There is an advantage to sending messages written in less common languages in two versions (original text in the body of the message plus attachment or converted text). Recipients whose e-mail clients support the given language can read the message in its original version, without having to bother with opening the attachment or converting text from Latin, while those whose software cannot process non-Latin characters can still read the message.



Another approach is based on substituting unsupported characters with something that stands for the special characters. For example, these can be capital letters in place of special characters ("C" for "ç"), or similar-looking letters from other alphabets.

This approach works only when there is a common understanding for what characters such substitutes are meant to represent.

Technical issues: fonts and encoding

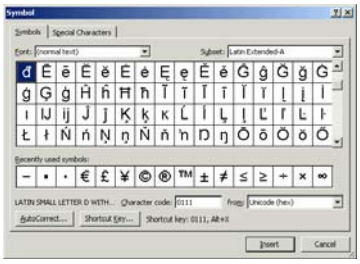
Typing in non-Latin characters

Even if you have the right software installed, typing your text into the computer can still pose **challenges if the keyboard you are using doesn't include all the characters used in your alphabet**.

English keyboards can be customized for many alphabets by labelling corresponding keys with non-Latin letters. However, alphabets that aren't phonetic or that use hundreds of symbols (such as Chinese) are difficult to type using conventional keyboards.

On-screen keyboards are a commonly-used solution to this problem. Their limitation is that although they can provide an infinite number of characters, selecting letters one-by-one using a mouse is cumbersome and takes significantly more time than when you use a real keyboard.

Most word processing programs include an "insert symbol" menu option which allows you to insert symbols and special characters not available on your keyboard.



"Symbols" options in MS Word

Technical issues: fonts and encoding

Web sites using non-Latin characters

If your Web site is primarily in a language which uses non-Latin characters you will need a thorough understanding of **fonts and encoding issues**.

Here is the home page of the United Nations' site in Arabic: http://www.un.org/arabic/.

If your Web site is in a language which uses Latin characters, but you occasionally need to provide content which uses other character sets consider one of the following options:

- make **content available via downloadable documents** such as PDF files or MS Word or OpenOffice.org documents; and
- **use images to display characters** that can not be effectively encoded as HTML.

Would you like to know more about fonts and encoding issues?
See Annex 4.6.1 for a mini-lesson on character encoding and encoding schemes.

Technical issues: fonts and encoding

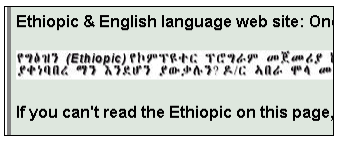
Use images to display characters that can't be effectively encoded as HTML

Capture all or part of the text as an image file, and display that instead of an HTML page.

This solution works well (and is usually the only one available) for texts containing alphabets that simply cannot be encoded, but such pages are slow to load, text is harder to read than conventional HTML text and it can not be copied.

It is also difficult to edit such page – a new image must be created from the original text every time (whether that is handwritten or word processed).

Below you can see an example of **"Image file" solution**: text is turned into an image file (.jpg or .gif) and displayed instead of regular HTML generated text.



Technical issues: fonts and encoding

As opposed to members of small language communities, English speakers don't need to deal with multilingual computing issues.

- True.
- False.

Please click on the answer of your choice.

Communication issues

When communicating in a foreign language it is more difficult to express ideas precisely. Language barriers can prevent you from sharing your knowledge as effectively as people who are communicating in their first language.

- When communicating in a **language that is not your own language**, expect that:
  - communication will be more demanding in terms of time and effort.
- When communicating in **your own language with speakers of other languages**, be prepared to:
  - simplify the language you use and avoid slang and colloquial expressions.
- When communicating in a **multilingual context** be prepared to:
  - face misunderstandings caused by different levels of knowledge of the language of communication and also misunderstandings caused by cultural differences; and
  - adopt communication practices which might differ from your usual cultural code of behaviour (for example, using language which is less or more formal).

Communication issues

Even if you can speak some of the dominant languages, it is good to use and promote software which is **localized for your language and region**.

You will then be able to communicate with people from your linguistic community who don't speak languages other than their mother tongue.

By promoting computing and Internet communication in local languages, you increase the chance that computers can be used by people who would otherwise be marginalized from the benefits of online communication.

Communication issues

Localizing software into a rich variety of languages contrasts language uniformity and distortion.

Current computers' inability to deal with the full range of languages leads to situations where people from certain language groups are forced to write documents and e-mails in typescripts other than their own.

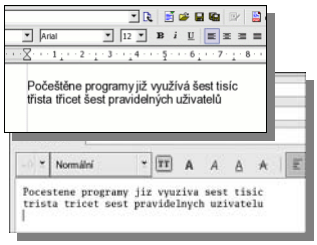
Communication issues

Writing documents and e-mails in typescripts other than mother tongue language

For example, some Slavic languages use a Latin character set plus extra accented characters for writing.

For e-mail communication, computer users from these linguistic groups commonly replace characters containing added special accents (such as "č"), by characters without accents ("c").

Another example is computing by the Arabic language community.



comparison of paragraph written in a text editor and an e-mail message

Such documents, though intelligible to readers, are grammatically incorrect in their language.

Such compromises may make communication possible, but they also mean that the languages concerned cannot be written correctly, which can serve to undermine the language. Only full software localization based on proper standards can enable all users to express themselves fully and correctly without running the risk that their message won't be readable.

Developing a language strategy

If your online community crosses language groups, it is important to **develop a language strategy** which promotes effective participation by community members, while at the same time **taking into account financial, time and human resource constraints**.

Your language strategy should cover all areas of your community's computing and communication needs: mailing lists, design and management of Web sites, e-mail communication, and documentation.

Options include:

1. **complete parallelism**, in which there are separate comprehensive newsletter, Web sites etc. for each language group, all with exactly the same content;

2. agreeing on a **single language of communication** for the group;

3. agreeing on a **primary language of communication but translating key documents into other languages** etc.

Developing a language strategy

Developing a strategy requires the following steps:

1. Mapping out linguistic context
2. Mapping out human and financial resources
3. Identifying a range of possible strategies
4. Choosing the most appropriate strategy

Let's look at these one at a time…

Developing a language strategy

Step 1 - Mapping out the linguistic context

1. Mapping out linguistic context
2. Mapping out human and financial resources
3. Identifying a range of possible strategies
4. Choosing the most appropriate strategy

Start by mapping out the linguistic context and needs of your community, answering the following questions:

- What are the primary languages spoken by community members?
- What other languages can they communicate in, and how effectively?
- Are there any political sensitivities relating to the use or exclusion of particular languages?
- Is there software to support the language/s concerned?

Developing a language strategy

Step 2 - Mapping out the human and financial resources

1. Mapping out linguistic context
2. Mapping out human and financial resources
3. Identifying a range of possible strategies
4. Choosing the most appropriate strategy

Multilingual online communities can require, for example, different facilitators for different language groups, or ongoing translation, or access to specialized software.

Think about the resources you have available to address these needs:

• Do you have volunteer facilitators or translators available?

• On an ongoing basis, or only for a fixed-time event?

• If you need to pay translators or buy software, what budget will you have available for this?

Developing a language strategy

Step 3 - Identifying a range of possible strategies

1. Mapping out linguistic context
2. Mapping out human and financial resources
3. Identifying a range of possible strategies
4. Choosing the most appropriate strategy

For example, if your community is made up of English, French and Spanish speakers, some possible strategies are:

1. to post information in only one of these languages (NOT recommended as it could cause problems for the community members who do not speak that language);
2. to post information in one language only, but to translate the summary in the other languages (a good option);
3. to post information in all languages (the best option, but requires resources for translation).

Challenges to be taken into account when designing multilingual Web sites

Developing a language strategy

Step 4 - Choosing the most appropriate strategy

Choose a strategy which balances community communication needs with the human and financial resources available.

1. Mapping out linguistic context
2. Mapping out human and financial resources
3. Identifying a range of possible strategies
4. Choosing the most appropriate strategy

Language strategies usually require compromises.

Ideally, you might want all community members to be able to participate in the community using their first language. However, few online communities have the resources to make this possible.

Look at your community's context and needs, at the resources you have available, and **decide on the best compromise for your group**.

Avoid choosing too ambitious approaches to multilingual issue. For example, accept that it might not be possible to duplicate complete Web sites in each language or run all online conferences in parallel languages.

An example of multilingual online community: the APC

Developing a language strategy

Put in the right order the main steps to follow in developing a language strategy for your online community.

1. Identify a range of possible strategies
2. Map out human and financial resources
3. Choose the most appropriate strategy
4. Map out linguistic context

Please order these items using the dropdown boxes and press "Check Answer".

Implementing the strategy

A key element in implementing the strategy is **communicating the strategy:**

- **discuss the proposed strategy with community members** before turning it into a formal policy, and be prepared to revise the strategy in line with community input;
- **explain your reasons for choosing to use a particular language or languages**; and
- if you are expecting many community members to use a language which is not their own, **ensure that they know why their participation in multilingual communities can benefit them** (for example by enabling them to network with people from other regions with the same interests).

Implementing the strategy

Once the strategy has been accepted, write up a short document which sets out:

- your online community's **language policy** (what information will be shared in which languages, how it will be archived etc.); and
- some **guidelines for effective multilingual communication**.

Be aware that working in a multilingual community requires **ongoing attention to language issues**. You should constantly try to:

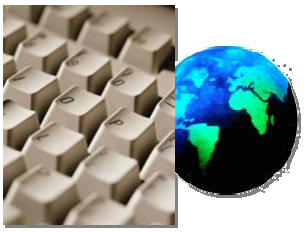
- **facilitate discussions** in such a way that the maximum number of participants can understand, and become actively involved (take care in your own writing, and encourage others in the community to express themselves in language which is accessible to other community members);
- look out for members who may not be participating because they cannot express themselves easily in the primary language of the community, and see if you can find ways to assist them.

Implementing the strategy

Writing for multilingual audiences

- Choose simple words.
- Keep sentences short.
- Choose more "formal" vocabulary (rather than using slang and other informal language which might be unfamiliar to your audience).
- Provide an explanation or definition in brackets if the word or expression is not common.
- Use the "active voice" when you can (e.g. "we sent her a message", rather than "a message was sent to her").
- Use less complex grammatical structures where possible.
- Define abbreviations and acronyms the first time they are used - for example, APC (Association for Progressive Communications).
- Avoid using phrasal verbs, where possible. A phrasal verb is a verb made up of two parts (verb + adverb) – for example "put off" (delay) and "put up" (provide accommodation for). Phrasal verbs can be a stumbling block for non-English speakers.

Machine translators



There are some useful translation tools that may facilitate multilingual communication:

- **free online translators**; and
- **standalone translation software**.

As well as these tools, you can also consider using volunteer services. In fact, some volunteer programmes also provide **volunteer translators** (see http://onlinevolunteering.org)

Would you like to know more about Online and Standalone Translators?
See Annex 4.6.2 for a mini-lesson on what they are and how they work

Machine translators

Getting the most out of machine translation

Machine translation works best within fairly narrowly defined contexts, using a "controlled vocabulary".

Defined contexts using limited and clearly defined terminology (for example specific areas of health or computing) are better subjects for machine translation. For example, when a machine translation system knows that the text to be translated concerns banking, a "card" is probably a credit card, not a jack of diamonds.

Use controlled language. You can improve the output of machine-processed translations significantly if you construct the text to be translated carefully. The simpler it is, the higher the accuracy of the translation:

•keep sentences short and simple;
•restrict the range of vocabulary to the most commonly-used expressions; and
•write in the active mode (e.g. "*they selected him*", instead of "*he has been selected*").

If you are running a Web site for an international audience and it is likely that your users will use machine translators for reading your web pages, apply the same rules to writing content for your Web site.

Summary

- Challenges relating to multilingual online communication relate to both technical issues such as software and fonts, and to human communication issues.
- Small language communities and communities using non-Latin character sets may face many barriers to computer use.
- Communicating in multilingual groups requires extra care and effort on the part of participants and facilitators.
- Localizing software can help overcome the digital divide by giving linguistically marginalized communities access to technology; open source provides good opportunities for translation and localization.
- Consult the community when developing the language strategy, and communicate the final policy to existing and new community members.
- Online translation tools may be of use in getting the general sense of a text in a language you are unfamiliar with, but are not a substitute for human translators.

If you want to learn more...

Online resources

ItrainOnline Multilingual Computing section
http://www.itrainonline.org/itrainonline/english/multilingual.shtml

United Nations Online Volunteering Program.
http://onlinevolunteering.org/

Eurescom. Guidelines for building multilingual Web sites
http://www.eurescom.de/public/projectresults/P900-series/923d1.asp

APC. Writing English Texts for a Multilingual Audience [also included as PDF file here]
http://www.apc.org/apps/img_upload/4f706479c23d02a8fd283c4020468eaa/multilingual.pdf

On-line translators
http://www.omniglot.com/links/translation.htm

Translation.net
http://www.translation.net/

IBM. Preparing your Web site for machine translation.
http://www-106.ibm.com/developerworks/web/library/us-mt/#7

Upgrade. Efficient Management of Multilingual Electronic Conferences.
http://www.upgrade-cepis.org/issues/2002/1/up3-1Pimienta.pdf

Kyrilcho. Latin Cyrillic on-line converter.
http://www.uni-bonn.de/~manfear/cyrlatencoder.php

The Knotty Problem of Using African Languages for E-mail and Internet Balancing Act
http://www.balancingact-africa.com/news/back/balancing-act_69.html

Translate.org.za
http://www.translate.org.za/

Khmer Software Initiative.
http://www.khmeros.info/

Harkus, S. 2001. Writing for machine translation
http://www.multilingualwebmaster.com/library/writing-TR.html

Annex 4.6.1
Mini-lesson: Character encoding in non-Latin scripts

Character encoding

It has probably happened to you, at some time or another, that you open a web page and find unreadable text and meaningless characters. This is a problem connected to **character encoding**.

Character encoding is the organization of a set of **numeric codes** that represent all meaningful characters (single letter, digit, space, punctuation, etc.) of a script system in memory.

Each **character** is **stored** in memory **as a number**. When a user enters characters, the user's keypresses are converted to character codes; when the characters are displayed on screen, the character codes are converted to the glyphs of a font.

A = 65

Character encoding

In most character encoding standards, the character set changes to represent the language being used, so the upper-level characters may include symbols, accented Roman letters, Cyrillic, or other characters, **depending on the character encoding chosen**.

For example, the character "Ó" in the Macintosh Standard Roman Character Set is in the same code point 205 as the equal sign "=" in Windows extended ASCII encoding.

Encoding schemes

Several encoding systems are available for which encoding schemes have been developed:

	7-BIT ENCODING SYSTEM	8-BIT ENCODING SYSTEM	16-BIT ENCODING SYSTEM	32-BIT ENCODING SYSTEM
What is	An encoding system that uses a fixed width of 7-bit encoding that allows for a character set of **128** values (2^7).	An encoding system that uses the eighth bit (parity bit) of the 7-bit encoding system to cover a larger number of characters. It allows for the use of **256** values (2^8).	An encoding system that uses a fixed-width of 16 bits per character, which allows the accommodation of a total of **65536** (2^{16}) values.	Standard named ISO/IEC 10646-1. It is essentially a 31-bit encoding, i.e., 2^{31} = **2147483648** code positions.
Schemes	**ASCII** and **ISO 646** are examples of 7-bit encoding. In fact, only **English**, **Latin**, and **Swahili** languages can use plain 7-bit ASCII with no additional characters. Most languages based on the Latin alphabet require a larger code set.	It covers **most common European languages**, such as French or German, that have accented letters, as well as **Arabic** and **Hebrew**. Many national variants were developed. To normalize the mess of 8-bit encodings, ISO came up with the **ISO 8859** series of standards.	It is needed for **Asian languages**, such as Chinese and Japanese that use ideographs, or hieroglyphs, instead of letters. **Windows NT**, for example, uses **16-bit internally** for all character values.	This system, also known as Universal Multiple-Octet Coded Character Set (**UCS**), was developed as standard in 1993. Today, most PCs have 32-bit registers.

Register sizes are rapidly growing to 64 bits. Special codes are now written for the 64-bit chip used in Windows XP.

Encoding schemes

The **encoding scheme** (also known as **character set**) represents characters used in a **different language**.

Encoding schemes

Using the charset code, you can **insert**, **edit** or **update** text in an HTML page **in the original language** of that page.

The charset code must be included in your HTML page by inserting some **META tags** in the **HEAD** section of the HTML page.

For example:

Employment
Emploi
Empleo

English, French, Spanish:
<meta http-equiv="Content-Type" content="text/html;**charset=ISO-8859-1**">

就业机会

Chinese:
<meta http-equiv="Content-Type" content="text/html; **charset=gb2312**">

فرص العمل

Arabic:
<meta http-equiv="Content-Type" content="text/html;**charset=windows-1256**">

Encoding schemes

You can also use UTF-8 (Unicode Transformation Format, 8-bit encoding form) as encoding, especially when you have to mix languages freely on the same page.

UTF-8 is an encoding form of the Unicode Standard, the universal character encoding standard used for representation of text for computer processing.

The Unicode Standard provides the capacity to encode **all of the characters used** for the written languages of the world.

One disadvantage of Unicode is that it takes more space to store plain text and transmission of Unicode data can therefore use more bandwidth.

Unicode is now widely used and has become the preferred character set for the Internet, especially for developing, processing and exchanging multilingual HTML and XML documents, and it is also being adopted for use in e-mail.

However, Unicode is not the most common character set in use. According to Microsoft, there is "built-in" support for Unicode in Windows NT and Windows 2000, but only limited support in Windows 95 and Windows 98. In addition, older office automation software such as Office 97, do not offer Unicode options.

Annex 4.6.2
Mini-lesson: Online Translators

Online dictionaries

A very primitive form of online translation is offered by **online dictionaries**.

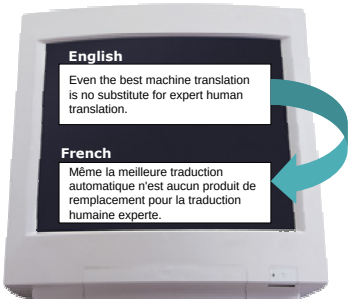
Online dictionaries are used in the same way as regular printed dictionaries, but instead of browsing through paper pages, you type the word you need to translate into an online form field.

There are online dictionaries for translating between multitudes of languages available on the Internet.

To find a dictionary for your language, use an Internet search engine (such as *Google™ search*). Your search could look like this:

Machine translators

Unlike dictionaries, which translate word-for-word, machine translators convert a *string* (phrase or sentence) in one language into a string in another language, using a limited knowledge of word order and grammatical rules of both languages.



Although machine translation has improved dramatically over the past few years it is still far from perfect.

Even the best machine translation is no substitute for expert human translation.

However, **it can be used to supplement human translation**, or it may be the only translation option available to you – so it is worth understanding machine translation basics and how to make the most effective use of the tools which are available.

There are two types of machine translators:

- **online translators**, are usually free and allow you to translate text from your web browser (or by e-mail); and
- **standalone machine translators**, are installed on your computer and used without being connected to the Internet.

Online translators

Online translators offer a useful if limited service. They allow you to paste in a piece of text and/or a URL, which is translated by a program hosted on the web server.

They commonly provide several language pairs (the languages **from** and **to** which you can translate – for example English/French, German/Spanish), and are usually free.

One of the most widely used free online translators is **Babel Fish**, a service provided from **AltaVista**.

Other examples are **Prompt-online**, **WorldLingo** and **Freetranslation**.

Ajeeb is a free translator between Arabic and English.

http://babelfish.altavista.com/

T-mail is a combination of free web-mail service and online translator. The service allows you to compose and create e-mails online as with other web-mail services, however, you can chose a language to which your e-mail message is translated, prior to sending.

Online translators

STRENGTHS

- They enable users to get the general sense of a text or web page in a language they are unfamiliar with. Although imperfect, this may be the only option – and better than no translation at all.

WEAKNESSES

- Free online tools provide the most basic form of machine translation. None of the options which allow for more accurate translation (for example specialized vocabularies) are available in these tools. These tools also limit the amount of text which can be translated at one time: for example, a single web page, or a limited number of words.

The **quality of translated content** you can get from a free online translator depends greatly on the combination of languages from and to which you are translating. Generally, translating between languages with similar grammar and syntax (e.g. Spanish and French) will give you much better results than translations between grammatically distant languages (e.g. Japanese and English).

Standalone machine translators

Like online translators, **standalone machine translation tools** convert strings in one language into another. Standalone commercial applications are significantly **more accurate** than free online translators.

For example, they allow users to add terms to the vocabulary database, and generally include some sort of "translation memory" database of previously translated content. The translators can either "learn" automatically or they can be "taught" - the user indicates which translating operations are correct and they are used in future.

Their prices range from about $50 to thousands of dollars, depending on the sophistication of the system. Some of these tools are designed as plug-ins to other applications (word processors, e-mail clients etc.).