8

Machine translation quality evaluation systems

8.1. MACHINE TRANSLATION QUALITY EVALUATION SYSTEMS

In recent times, there have been numerous attempts to develop criteria for assessing the quality of text information processing systems for natural languages. But still there is no single evaluation procedure. However, there are many professional developers and users interested in developing evaluation criteria that are compatible with different operating characteristics of current systems.

Evaluation is a key component in any technology. There is a need to assess the general characteristics of systems. A characteristic can be measured for different developmental stages of the same system. This kind of comparative assessment will be called vertical. Alternatively, the quality of multiple systems can be compared at a given time. This second type of comparative assessment we will call horizontal. Vertical assessment refers to the product developers of text processing systems and servicing.

Horizontal assessment refers both to developers and end users. For developers - to determine the relative quality of systems compared with systems of their competitors and to potential end-users to make an informed conclusion about what products can best meet their requirements.

In recent years, interest in the theoretical and practical research on machine translation (MT) have been growing rapidly. MT depending on the technical means can be mechanized, automated and automatic. Indeed, the mechanized system provides some technical means that enhance information processing. Automated system involves the processing of information based on well normalized actions with use of computers and other technology that ensures minimal use of capabilities of human and machine. The term "automatic" usually refers to any process or procedure that performed without human intervention.

In this regard, the existing methods of machine translation - lexical, grammatical and semantic - are the stages of its automation. Implementation of all existing techniques into a single system allows for only an automated and/or semi-automatic processing of information in a bilingual situation. Fully automatic system is usually thought of only when the implementation of audiovisual input and output of information from the original source, without human intervention, hence the term "automatic translation" in its full and literal sense characterizes the highest and has not yet reached stage of machinery in the translation system.

8.2. CRITERIA AND METHODS FOR ASSESSING THE MACHINE TRANSLATION QUALITY

Assessment of the quality of text processing systems should be carried out by the following criteria: quality of software; user interface usability; linguistic description, organizational value, price efficiency. Let us describe some of them in detail.

System Specifications

- ⚘ speed and translation delivery time
- ⚘ editor features
- ⚘ dictionaries (possibility and ease of updating)
- ⚘ fault tolerance
- ⚘ usability of information
- ⚘ value

Criteria for evaluating the output text

⚔   as a comparison of the characteristics of the original meaning and translation:

translation accuracy

correct terminology translation

⚔   as characteristics of the text content:

intelligibility

clearness

⚔   As the characteristics of representation in the target language:

naturalness

linguistic correctness

coherence

Methods of assessing the quality of output text:

⚔   back translation

⚔   comparison with the reference translation

⚔   transformation method

⚔   denotative method

⚔   user surveys

⚔   error analysis on the linguistic level (vocabulary, grammar)

Procedure for evaluating the software package reliability should be performed on the increase. The first step is to evaluate the reliability of software modules. As the initial data for this stage it is better to use the figures of fault-free execution of machine instructions and operators of high-level languages cited in the literature. These indicators are usually obtained by processing large volumes of statistical data and confirmed by practice. Estimation of probabilities of the branches of a graph or control transfers between the modules can be done by analyzing the original data by an expert, or the realization of the equal probability hypothesis of all paths from each branch point. Using the latter option leads to the fact that the assessment of reliability is low, and in this case it has the maximum error.

Evaluation of the reliability of software system can only be done after finding the parameters for all components included in its composition, and should be implemented along the same lines as the module performance assessment.

The very nature of natural language prevents from using completely objective and formal assessment methods. Regardless of how computing and automated method will be, it will always have a proportion of subjectivity. It is necessary to minimize the proportion of subjectivity and the amount of human effort associated with the development of a method to find practical and financially affordable decision.

8.3. REVIEW OF ONLINE TEXT TRANSLATION SYSTEMS

With developing of new methods and software solutions for machine translation, the global market gets filled with lots of programs. Today the Internet offers more than a hundred services, providing on-line translation. Many companies are engaged in development of machine translation software today, but products of the two organizations are leading the world market - Systran and Russian company PROMPT. Other major producers include Linguatec and Langenscheidt (Germany), Transparent Language, Babylon, Translation Experts (USA), Asia Online Sino-Thai company, etc.

According to the rating of existing on-line translation services and views of users of various forums devoted to modern on-line translators, we have identified the most popular online services available to the modern Internet users:

1. Translate.ru (online PROMT translator)
2. SYSTRANet (Systran)
3. Google Trnslate (Google)
4. Free Translation (SDL)
5. Babel Fish (Systran)
6. Worldlingo (Systran)
7. InterTran (Translation Experts Limited)
8. ImTransator (Smartlink Corp)
9. Tranlslate Online.ua
10. Windows Life Translator (Microsoft)

Many software developers offer online translation services as an advertising for their professional translation software.

So, let's consider the features of each service online translation.

1. Translate.ru (PROMT)

http://www.translate.ru

PROMT is the absolute leader on the Russian market of online translation. This service has existed since 1998 (for foreign users there are versions in other languages at: www.online-translator.com).

PROMT Online Translator supports seven European languages (English, Russian, German, Italian, Spanish, Portuguese and French) and provides translation services in 25 directions:

from English — into Spanish, Portuguese, Italian, Russian, German and French;

from Spanish - into English, German, French and Russian;

from Italian - into English and Russian;

from German - into Russian, English, French and Spanish;

from Portuguese - into English;

from Russian - into Spanish, English, French and German;

from French - into Spanish, English, German and Russian.

When translating specialized texts, you can choose a theme - communication, business, computers, cars, travel, sports, or "other." In addition, by registering for free on the site, a user can select additional options (connect specialized dictionaries, install the settings for large text translation, get grammatical information). The disadvantage of the program is then limit on the amount of input text up to 500 characters.

In addition to translation, this online translator also offers web-pages translation services with full preservation of the original formatting, electronic communications (e-mail messages), as well as WAP (translation of individual words for mobile users). If by clicking on the appropriate link you want to get translation of a web page, simply select "Automatic links translation" option, choose a topic from the list and click on "Translate".

Among the additional amenities: a virtual keyboard, spell checking capabilities, optional setting of unknown words transliteration and the conversion of Cyrillic. In addition, there are various opportunities to work with translation - it can be e-mailed, printed or simply copied from the field for the translation.

2. SYSTRANet

http://www.systranet.com/

Online translation tool offered by another world leader in the field of machine translation - the american company Systran.

Online text translation service supports 15 languages: Arabic, Chinese, English, Dutch, German, French, Italian, Greek, Korean, Japanese, Portuguese, Russian, Polish, Swedish and Spanish. In total, this translator provides translations in 53 language directions:

from Arabic, Chinese, Japanese, Polish, Korean, Russian and Swedish languages - into English;

from Chinese – into English;

from Dutch - into English and French;

from English - into Arabic, Chinese, French, Dutch, Greek, German, Korean, Japanese, Italian, Polish, Russian, Portuguese, Swedish and Spanish;

from French - into Dutch, German, Greek, English, Italian, Portuguese and Spanish;

from German - into Italian, Portuguese, English, French and Spanish;

from Greek - into English and French;

from Italian - into German, English, French, Spanish and Portuguese;

from Portuguese - into English, German, Italian, French and Spanish;

from Spanish - into Italian, English, Portuguese, German and French;

Besides usual functions of translation SYSTRANet also provides translations of Web pages and files with following extensions: .txt, .htm, .rtf. In addition, the user is prompted to select optional subject dictionaries or create own dictionary. But in order to create own dictionary or to upload a file in supported format, a user must first register on this site.

3. Google Translate

http://translate.google.com/
http://translate.google.ru/

Google Translate service is an online translator of texts and web pages, which is a part of Google search engine. This online translator works with 51 languages (English, Afrikaans, Arabic, Albanian, Bulgarian, Byelorussian, Welsh, Vietnamese, Hungarian, Greek, Dutch, Galician, Hebrew, Danish, Yiddish, Indonesian, Irish, Icelandic, Spanish, Italian, Korean , Catalan, Chinese, Lithuanian, Latvian, Malay, Macedonian, Maltese, Norwegian, German, Polish, Portuguese, Persian, Russian, Romanian, Swahili, Slovak, Serbian, Slovenian, Thai, Turkish, Tagalog, Ukrainian, French, Finnish, Croatian, Hindi, Czech, Swedish, Estonian and Japanese) and, in general, perform translation in 2250 directions. This service has several new convenience features. Now there is no need to use "translate" button, because the program automatically translates a word or text as it is typed in the box. When translating individual words, the application operates as an online dictionary and offers all possible translations, breaking all the words by parts of speech. In addition, when translating jawbreaker words program prompts the user to select the option "Show transliteration", which carries spelling of the word with extended Latin alphabet. This option is available for all languages that use non-Latin alphabet, with the exception of Arabic, Persian and Hebrew. Like other online text translator, Google translates web pages. In addition, it supports an additional search function "Translated Search, which allows to find sites in different language with translation of a word or a sentence.

Perhaps in functionality - the number of supported languages and areas of additional features - online Google translator is the most versatile among the other online translators. However, we have to learn, if it is strong translator in question of the quality of translation.

4. Free Translation

www.freetranslation.com

Online translation tool Free Translation, provided by SDL, supports 32 languages (not counting some regional variants of languages): Arabic, English, Bulgarian, Chinese (simplified and traditional versions), Dutch, Danish, French, Finnish, German, Hebrew, Greek , Hindi, Hausa, Italian, Hungarian, Korean, Japanese, Persian, Norwegian, Portuguese (Brazilian and European versions), Polish, Pashto, Romanian, Russian, Somali, Spanish (Latin American, Mexican and European versions), Serbian, Czech, Swedish , Thai, Turkish, Urdu. In total, this translator provides translations in 86 language directions. Most of the translations done from English, but at present translation services for some languages (eg, Turkish, Thai, Urdu) are unavailable for unknown reasons. Online Translator Free Translation also performs translation of web-pages.

Unlike other online translation of the text using Free Translation provides additional options to improve the quality of translations to exclusive members of FreeTranslation Platinum Club, registration in which costs a fee of 9.95 dollars (per year). Additional options allow usage of specialized dictionaries in different areas: Business, IT, Current Affairs - and translate large texts up to 10 thousand characters.

In addition to machine translation services, Free Translation also offers paid professional translation services by qualified interpreters.

5. Babel Fish

http://babelfish.yahoo.com/

One of the first computer-aided translation systems available online was launched in 1997 by AltaVista, but was later integrated into the Yahoo search engine, complemented this online translation service with new languages. Babel Fish Online translator, as well as the next translator Worldlingo, uses Systran translation technology. On-line "fish" translates texts up to 150 words and Web sites in 38 language directions:

from Simplified Chinese – into English and Traditional Chinese;

from Chinese traditional - into English and Simplified Chinese;

from English - into Chinese Simplified, Chinese Traditional, French, Dutch, Greek, German, Italian, Korean, Japanese, Portuguese, Spanish, Russian,

from Dutch - into English and French;

from French - into English, Dutch, German, Italian, Greek, Spanish and Portuguese;

from German - into French and English;

from Greek - into English and French;

from Italian - into English and French;

from Japanese, Korean and Russian languages - into English

from Portuguese - into English and French;

from Spanish - into English and French

Babel Fish limits the amount of text by 5 Kbytes (approximately 800 words), this constraint is preserved when translating web-pages.

As a rule, quality translations performed by Babel Fish, is comparable to the quality of online translations from Google Translate, but inferior to the quality of translations from Translate.ru. When translating this program typically loses a breakdown into paragraphs.
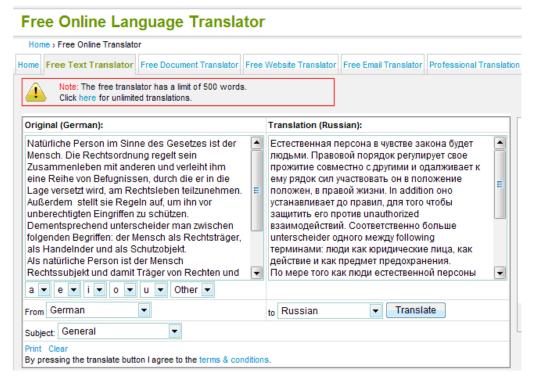

6. Worldlingo

http://www.worldlingo.com/en/products_services/worldlingo_translator.html

This online translator supports 32 languages (Arabic, English, Bulgarian, Danish, Dutch, Farsi, Finnish, French, German, Greek, Czech, Hausa, Hebrew, Hindi, Hungarian, Italian, Spanish, Japanese, Korean, Chinese (simplified and Traditional options), Norwegian, Pashto, Polish, Portuguese, Russian, Romanian, Serbian, Swedish, Thai, Turkish, Urdu) and allows to translate texts up to 500 words online. This online translation has the advantage in terms of features - the ability to select the 20 specialized subjects and special characters specific to each language.

In addition to translation of websites and texts this online translator can translate text files and emails. To translate e-mails, user must specify the name and email address of the sender and recipient, specify the language direction, if desired, set the theme and add a text message, if necessary, using special characters. After this

translated message is sent to the relevant addressee at entered address. If for translation of websites, emails and texts there is a limit on the volume, then for files translation, user can upload documents of any size. Existing limitations can be removed for a fee.

7. InterTran
http://www.tranexp.com:2000/Translate/result.shtml

InterTran online translator (InteractiveTran) represents one of Translation Experts Limited company's products. Translator supports 27 languages (English, Danish, Bulgarian, Dutch, Czech, Spanish (and Latin American), French, Finnish, Greek, German, Hungarian, Icelandic, Italian, Japanese, Norwegian, Philippines (Tagalog), Polish, Portuguese (as Brazilian), Russian, Romanian, Slovenian, Serbian, Swedish, Croatian, Welsh, Turkish and Latin). Text for translation can be entered in any of these

languages, but the disadvantage of translator is that it takes only the text and does not allow to select specialized dictionaries while translating the text of certain subjects.

8. Im Translator

http://translation.imtranslator.net

Online service Im Translator is a product of the Smartlink Corp american corporation, based on database technology developed by Russian company PROMPT. Unlike Translate.ru, Im Translator comes with a standard set of features and performs a translation in 19 language directions:

from English – into German, French, Russian, Spanish and Portuguese;

from French - into Russian, English and Spanish;

from German - into Russian and English;

from Portuguese - into English;

from Russian - into English, German, French and Spanish;

from Spanish - into French, English and Russian;

from Russian - into Turkish.

Additional functionality of the online translator are the availability of electronic dictionary and spell checker. However, the translator does not offer the user the option to connect the thematic dictionaries.

9. Translate Online.ua

http://pereklad.online.ua/

It's Ukrainian online translation service that supports 7 languages (Ukrainian, Russian, English, German, French, Latvian and Polish) and allows to translate texts with the restriction on the size of one kilobyte in 42 language directions.

Translator allows to choose the proper subject of a text: auto, business, internet, general, law and engineering. In addition, there is a simple but very useful feature - below the text entry field there is a small handy dictionary.

10. Windows Live Translator

http://www.microsofttranslator.com/Default.aspx

Microsoft has launched its online translation service Windows Live Translator. This online translator is virtually identical to the standard of online translators. Translator supports 19 languages: Arabic, English, Danish, Greek, Hebrew, Italian, Spanish, Korean, Chinese (Simplified and Traditional), Dutch, German, Portuguese, Polish, Russian, French, Thai, Swedish, Czech and Japanese. All translations from English into European languages (except Russian), Chinese and Japanese performed by translator, developed by Microsoft, and all other translations are based on Systran translation technology. Convenience of the online translator is a Russian-language interface is installed automatically, depending on the IP-adress. However, the developers have imposed some restrictions on the functionality of this service: in the translation mode user can upload a text only up to 500 characters.

Comparative table of programs for automatic translation of text

The main characteristics of all these online translators can be summarized in the following comparison table

Name translator	Functionality	Supported languages	Limitations
PROMT	Free online service, which provides services to automatically translate texts, web pages, e-mail and WAP Possible to choose a special theme	7 European languages 25 lines	Before 500 characters
SYSTRANet	Service free automatic translation of texts, web pages and files with extensions. txt,. htm,. rtf.	15 Asian and European languages, 53 languages lines	Before 3000 characters
Google Translate	Free service of automatic translation, the translation of texts, web pages and documents. The service is a transliteration of words The search function "Search with the transfer" (web pages)	Asian and European Languages 51 languages, 2250 lines	unlimited

SDL FreeTranslation	Free online translator of texts and web pages Additional options to improve the quality of translations for members FreeTranslation Platinum Club (9,95 USD of year) - the choice of specialized theme + the ability to translate texts of up to 10000 characters	Asian and European languages, 86 lines	Before 10000 characters
Babel Fish	Free online service for automatic translation of texts and web pages Lost split into paragraphs	Asian and European languages, 38 lines	Before 150 words
WorldLingo	Free translator performs online text translation, web pages, text files and emails. Choice of 20 specialized theme and special characters specific to each language.	Asian and European languages, 32 lines	Before 500 words
InterTran	Free service automatically translate texts Not to select a specialized dictionary for translating the specific theme	27 languages	Before 500 words
ImTranslator	Free online translator supports spell orthography and special characters, as well as the transliteration option. Does not suggest a connection of specialized dictionaries	19 languages lines	Unlimited amount of text

Translate Oline.ua	Free text translator Offers a choice of specialized theme Under the window I have dictionary	7 languages, 42 lines	1 Кб
Windows Live Translator	Text online translator provided by the beta. Convenient Russian interface	19 languages	Before 500 characters

8,4. RECOMMENDATIONS FOR IMPROVING THE QUALITY OF THE WORK WITH TRANSLATION

There are ways to improve the results of machine translation available to each user. Some of them we will describe bellow:

1. Before proceeding to the translation, be sure to specify two things: firstly, what is the intended purpose of its outcomes, and secondly, what the original text is.

Purpose of translation plays a pivotal role in the assessment of its quality. Indeed, the same result can be considered excellent if you just need to know what the original article was about, and totally unusable, if you want the text to be published in a book or magazine.

Determining to which style of speech the original text belongs, it is easy to assess its suitability for machine translation, and hence to predict the outcome. The more allusive expressions, metaphors, the freer style, the worse your computer cope with its transformation. Scientific, technical and educational texts characterized by a rigorous exposition of the material are handled better than others. If special dictionaries are filled in time with new terms, it is possible to get a fully coherent text translation, requiring a minimum stylistic refinement.

Conversational and journalistic style with many specific terms, but most of the words used in the literal sense, are suitable for the trial translation, but for correct output text it will require manual editing. Poetry and fiction, unfortunately, quite unsuitable for machine translation. Meaning of the text, which was built on the

figurative expressions is distorted with machine translation and does not suit even for understanding.

2. Grammatical mistakes in the text. Very often typos in the original are the cause of mistranslation. This applies particularly to the scanned and recognized text. Misspelled words in most cases are flagged by the system as unknown, as they are not in the dictionary.

It is more complicated if a misprint turns one word into another, which also exists in foreign language - the program will translate it, but the meaning of the text will be distorted. But the most serious pitfalls are errors in punctuation. One wrong comma can seriously distort the translation of a sentence. Therefore, before the translation the source text must be carefully checked.

3. Using of standard dictionary is not always sufficient. Be sure to find and connect the special dictionaries on the subject of the text. If there is no such specific subject, determine the most suitable combination of existing dictionaries and, in addition, be sure to create your own.

4. Building the "pyramid". If several dictionaries connected to the system, success of translation depends on the order in which the program searches current word in them. Therefore, organize the hierarchy of dictionaries in order from specific to common. The highest priority should be given to a dictionary, created for the current text, and then – specific for subject (in an extension of the domain), and the lowest is for dictionary of commonly used concepts.

5. Never translate the entire text at once. There always exists at least one or two words, not in the dictionary, and big set that the system translates wrong. Having dealt with the connection of dictionaries and the definition of other options of translation, start with a small snippet at the beginning of the text, for example, the first paragraph. Find the word translated incorrectly in this passage, and add them to the dictionary of the highest level. Turn the piece over again. If you are satisfied with the result, go to the next paragraph.

Practice shows that for fine-tuning of the system fourth, and sometimes even a third of the material must be translated so, and only then automatic processing of the text is possible.

6. Distinguish between common and specific! Before you add a new word in the custom dictionary, determine whether it will be translated in the text in this manner only once or twice or permanently. Art of proper work with machine translation system is to distinguish between common and specific. Ideally you will be required to know foreign language and subject area of the text.

The dictionary should include only regularly occurring variants of the translation, the rest should be fix manually. Otherwise, the word will be translated incorrectly throughout the text.

Compliance with the simple rules above generally can significantly improve quality of texts translated by computer. It is impossible to guarantee that they will always be absolutely correct and stylistically literate.

8.5. CONCLUSION

Results of the comparison led to definite conclusions about the functionality and suitability of modern online translation services. Analysis of the translations made with selected online translators, suggests that these services provide the user with low-quality translations with a lot of mistakes. Of course, in some cases a translater rather successfully cope with the communication problem - sometimes it is possible, even though intuitively, understand the meaning of the original text. Of course, machine translation programs often have no potential to detect and correctly analyze certain grammatical constructions or lexical items, resulting in an incorrect translation and, therefore, is becomes distorted and difficult to understand the translated text. Therefore, the only suitable option in this situation would be painstaking comparison of source and translated text on the subject of the omitted word translation and editing them, of course, if there is sufficient experience and appropriate skill level for that. Based on the analysis of the translations, we can conclude that some of the online translators do have the potential to enable them to recognize the grammatical structures and lexical items and correctly reflect all the syntactic and grammatical context in translation. A deeper analysis of the grammar translation shows that some online translators are pretty successfully cope with the transfer of isolated parts of speech and words in the plural and singular. However, we have seen that when there

is a need for translation of connected complex sentences or longer texts, online translator is not able to adequately cope with this task. In addition, the great disadvantage of online translators is the wrong choice of the translations of the polysemous words. All of these errors, which distort the meaning of the source text and hinder understanding of the translation, due to insufficient supply of electronic translator algorithms required for in-depth analysis of grammatical structures and their adequate interpretation in the translation.

Of course, we could see that not all translators provide translations of the same low quality, so it would be unfair to evaluate them by one criterion. Among the selected service providers for online translation it is possible to distinguish a pair of translators who most successfully coped with their task and properly convey the meaning of the original text, retaining almost all the grammatical and syntactical context in the sentences. These services include Google, Translate.ru (PROMPT). Despite the apparent error in the translation and the presence of errors in grammar and speech, these translators are more successfully managed with translation and were able to convey the original meaning of the text, unlike the others, who provided no coherent, literate text, and the usual gloss. These include InterTran, Windows Live Translator, Translate Online.ua, issuing low-grade texts, whose meaning is difficult to reach the reader because of gross grammatical errors and incorrect translation of terms. Online translators do not have the capabilities and baggage of theoretical knowledge, which has a live interpreter, so they can not replace them, as though the developers may have tried to persuade users to return. Whatever may have been enhanced in machine translation system, it can not compete with native speakers, therefore, for them is a daunting task to restore all the words omitted in the text, or understand the true meaning of jargon or colloquialisms. More or less acceptable results can be expected when translating from one language to another within a common kinship group, but often even the most advanced translators give a complete nonsense to translate various texts and websites. Even if the online translators and other services of machine translation will be provided with a strong vocabulary base and a sufficient amount of grammatical algorithms, capable of providing a decent quality of translation, before such translators still will be an unattainable goal - to send an emotional background and imagery of the original text, which can only be expected

from living, thinking translators who are able to transfer thoughts and feelings that have been embedded in the source code is not a machine, and the same human beings. Yet, despite the low quality of translation, modern on-line translators are quite common on the Internet, from which it can be concluded that users often resort to their help. Therefore it is more correct to objectively evaluate the actual state of development of systems of online translation and consider the obvious advantages and disadvantages of virtual translators. Of course, an undeniable advantage of online translators is their availability at any time and any place where there is access to a global network, in addition, users appreciate the high speed of translation. Of course, high speed translation is a big plus of on-line translators, especially if you are faced with the need to perform urgent translation of foreign text, but keep in mind that, most likely, you have to spend on editing the resulting translation almost all the time (and sometimes more), saved with online translator.

It is important to consider all these moments and remember that the electronic translator is not designed to replace live interpreters, and to help the participants to modern intercultural and multilingual information sharing to overcome language barriers due to lack of knowledge of foreign languages.

REFERENCES

Винокуров А.А.,Чуканов В.О.Новый метод оценки машинного перевода.// Информационные технологии и системы. Hardware Software Security. Тенденции и перспективы - Сборник статей: М., Международная академия информатизации, 1997г.

Мосты. Журнал переводчиков. №4. 2004г.

http://ikit.edu.sfu-kras.ru/files/12/ch01s01s05.pdf

http://www.technofresh.ru/technology/internet/online-translaters.html

http://alemeln.narod.ru/pages/progs/progtr1.html

http://whoyougle.ru/texts/online-translator/

http://www.primavista.ru/

http://alemeln.narod.ru/progper2.html

http://www.innovativeeducation.com/lhs/tips.asp

Rakhimova D.R.
Zhumanov Zh.M.
Al-Farabi KazNU, Almaty, Kazakhstan