2. В. Гергель: Современные языки и технологии параллельного программирования. М., Изд-во МГУ, 2012.

3. Синюк, В. Г. Алгоритмы и структуры данных. Белгород: Изд-во БГТУ им. В. Г. Шухова, 2013.

REALIZATION SORTING ALGORITHM USING PARALLEL TECHNOLOGIES

bachelor, Mikhelev Vladimir
candidate of Science, prof., Sinyuk Vasily
Shukhov Belgorod State Technological University

Annotation *A comparative analysis of serial and parallel algorithms for sorting, for each test algorithm determined the dependence of acceleration factors on the number of calculators.*

Keywords: *sorting algorithms, parallel technologies, OpenMP, sorting vesicle sorting exchange, QuickSort*

УДК 004.9+81'322.4

PECULIARITIES OF MACHINE TRANSLATION TECHNOLOGIES IMPLEMENTATION IN MAJOR ON-LINE TRANSLATORS: COMPARATIVE STUDY

Morel Morel Dmitry Alexander, Ph.D. in Philology, associate professor of the Foreign Business Language Department,
Gimazitdinov Evgenij Igorevich, student
Belgorod National Research University

Annotation *The article considers the main existing machine translation technologies and their application in on-line translators by PROMT, Google and Yandex as well as their advantages and disadvantages in terms of output texts quality,*

Key words: *machine translation, Rule-based Machine Translation, Statistical Machine Translation, Hybrid Machine Translation, Translation Memory, PROMT, Google Translate, Yandex.*

Machine translation (MT) has made drastic progress for few decades of its history since the first suggestion for using computers for translation was made in the USA in 1947 at the break of dawn of the Computer Age. Having its practical start in 1954 with experiments at Institute of fine mechanics and computing machinery, USSR (researchers of I.K. Bel'skaya and D.Yu. Panov) and at Georgetown University, USA, this branch turned out soon to be economically sound [8].

Nowadays, MT is a booming IT-sector with many companies (Systran,

PROMT, Linguatec, Google, Microsoft, IBM, Atril, and others) involved in with a great number of commercial products being rolled out. On-line translation has become its promising trend being implemented in Ectaco, Google Translate, ImTranslator, InterTran, Reverso, Translate.ru, Windows Live Translator, WorldLingo, Yandex.Perevod, and other projects.

Such a variety can pose some problems for an end user—especially for one who is not well grounded in informatics and linguistics—the most burning of them being the estimate of efficiency of one or another on-line translator.

An overview of currently used machine translation technologies and a comparative analysis of features of their implementation in particular projects both made in the present paper can help a better understanding of this problem.

To limit the scope of our research subject we confined ourselves to the examination of only three major on-line translation services largely used in Russia: Google Translate, PROMT, Yandex.Perevod.

We can start by comparing three technologies of machine translation:

– Rule-based Machine Translation (RBMT);

– Statistical Machine Translation (SMT);

– Hybrid Machine Translation (HMT).

RBMT technology relies on specific languages dictionary data and grammar rules analysis, involving linguistic descriptions of pairs of natural languages, formal grammars, and translation algorithms proper. Thus, the quality of translation correlates with both the extent of linguistic databases and the profundity of natural languages descriptions, which means that we should take into consideration as many peculiarities of grammar structure of both source and target languages as possible [2; 10].

There are two kinds of RBMT technologies: transfer-type—implying morphological, syntactic and semantic analysis of a source language text, transformation into a target language structure, text synthesis in a target language—and Interlingua-type ones analyzing an incoming text in terms of a metalanguage and synthesizing the metastructure of the text in a target language [9].

RBMT technologies have both advantages (syntactic and morphological accuracy, stable and predictable outcome, knowledge domain customizability) and disadvantages (labor- and time-consuming development, obligatory maintenance and updating of linguistic databases).

SMT technology relies on searching the most probable translation of a sentence comparing large parallel corpora—sets of texts in one language and of their analogs in another. The more such sets are available and in keeping the better is the result, thus this technology can be referred to as self-training [2; 7].

SMT technologies also have their advantages (fluency of translation, portability to any pairs of languages, ease of building in the presence of sufficient number of parallel corpora) and disadvantages (limitedness of existing parallel corpora in number; inability to handle correctly neither morphology nor syntax, misrepresentation—duplication, omission, substitution—of information).

HMT technology is based upon the combination of RBMT and SMT methods. Such an approach makes it possible to bring into play strengths of both ones: grammatical accuracy of RBMT and translation fluency of SMT [3].

Besides the above-mentioned MT technologies there are machine translators whose principle of operation is based on Translation Memory (TM) technology that in turn uses as basis the principle not to translate the same sentence twice. This technology is grounded on the comparison of a document to be translated with the data contained in a prebuilt database of translations. Translator finds out once translated sections in the whole text array and retrieves their existing translations from TM database [1; 6].

TM technologies have both advantages (reuse of once made translations, minimum post-editing needed) and disadvantages (advance data entry is required; reuse depends on the resemblance of the content to be translated and TM databases; translation of a brand new content is impossible).

Having overviewed the main MT features let's pass to the analysis of their implementation in three selected on-line-translators.

The major contribution to MT development in Russia was made by a research

group under the guidance of P.G. Piotrovsky that afterwards founded the Russian firm PROMT developing and promoting the first Russian commercial MT-application of the same name.

PROMT on-line translator used to be based on RBMT technology (see above) but since 2010 has seen a shift to HMT one that made it rapidly trainable. The system using such a converged technology generates a multitude of a single sentence translations—which number can reach several hundreds depending on polysemy and statistical processing results—instead of one. Then probabilistic model of language allows to retrieve the most probable variant among the proposed ones [3].

Google Translate was developed by Google in mid-2000s for the on-the-fly translation of texts and web-sites. This on-line translator relies on SMT-rules-based technology and uses a self-training MT algorithm based on linguistic text analysis. The main feature of Google Translate is its approach to translation process: unlike other translators this system does not analyze grammar rules and vocabularies but seeks for linguistic correspondences between a text to be translated and a huge array consisting of human-made translation samples and involves self-training statistical algorithms for building translation patterns. Such a method ensures the quality and credibility of an output text maximizing elimination of incongruous collocations typical for other MT systems [5].

Yandex.Perevod is a web-service of Yandex Company intended for translating texts or web-pages using self-training SMT algorithm. The system builds its own vocabularies of equivalents basing on the analysis of millions of translated texts. The text to be translated is first compared to the words database, then to the language patterns database as the system tries to define the meaning of an expression in context [4].

Yandex MT system has three main parts: translation pattern, language pattern, and decoder.

The translation pattern is a table containing all possible translations and their probabilities for all words and sentences known to the system. It is created in three stages: the selection of parallel texts, then the selection of pairs of sentences within

them, and finally the selection of pairs of words or collocations. The system compares not only single words but also n-grams. Yandex.Perevod translation pattern contains hundreds of millions of pairs of words or collocations for each pair of languages.

To build the language pattern the system analyzes hundreds of thousands of different texts in the necessary language and draws up a list of all words and collocations used there with their frequency producing the system's knowledge of a target language.

The decoder is engaged in translation proper by selecting all variants of translation for each sentence of an incoming text—combining sentences from the translation pattern with each other—and by putting them in frequency descending order. All produced variants are evaluated with the help of the language pattern, eventually the sentence with the best match of probability (in terms of translation pattern) and frequency (in terms of language pattern) is selected by the decoder.

To illustrate the practical outcome of implementing the above-mentioned technologies in the on-line translators under study we chose a fragment from an article:

"As technology improves and Artic sea ice melts, multiple nations are fixing their gaze on the natural resources in the Arctic Circle. Scientists project that the region encompasses 13% of the world's undiscovered oil reserves as well as 30% of the world's undiscovered natural gas reserves. If one nation is able to assert legal or practical dominance over the region and harvest even a fraction of those resources, that nation would secure its energy future for years to come while also creating an economic boom." (Weidinger, M. (2013) "Race for the Arctic: Russia's Greatest Journey in our Time", Matters of Russian and International Law, 2, pages 157-194.)

Variant of translation by PROMT:

"Когда технология улучшается, и морской лед Artic тает, многократные страны закрепляют свой пристальный взгляд на природных ресурсах в Северном Полярном Круге. Проект ученых, что область охватывает 13% неоткрытых запасов нефти в мире, а также 30% неоткрытых запасов

природного газа в мире. Если одна страна в состоянии утверждать юридическое или практическое господство над областью и получить даже часть тех ресурсов, та страна обеспечила бы свое энергетическое будущее в течение многих последующих лет, также создавая экономический бум."

Variant of translation by Google Translate:

"Поскольку технология улучшает и Artic морской лед тает, несколько народы фиксации свои взоры на природные ресурсы в Северном Полярном круге. По прогнозам ученых, в регионе охватывает 13% неразведанных запасов нефти в мире, а также 30% неразведанных запасов природного газа в мире. Если одна нация способна утверждать правовых или практических господство над регионом и урожай даже часть этих ресурсов, что страна будет обеспечить свое энергетическое будущее на долгие годы в то же время создавая экономический бум."

Variant of translation by Yandex.Perevod:

"Как технология улучшает и Арктический морской лед тает, несколько Наций исправлять свои взоры на природные ресурсы в пределах Северного полярного Круга. По прогнозам ученых, в том, что регион охватывает 13% неразведанных мировых запасов нефти и 30% неразведанных мировых запасов природного газа. Если одна нация не в состоянии утверждать правовые или практические господство над регионом и урожай даже часть тех ресурсов, что нация могла бы его безопасного энергетического будущего на годы вперед, а также создание экономического бума."

As we could see all the produced translations are somewhat different in forms of rendering the incoming text content and manifest deviations from the norms of the target (Russian) language.

In conclusion we are to point out the following.

1. Machine translation technologies have both advantages (they can be used on any content and have high translation speed) and disadvantages (their efficiency depends on obligatory presettings and initial text quality). Such weaknesses of both SMT and RBMT as well as reached saturation point in their progress caused

developers to seek a solution through convergence. Thus, a technological breakthrough is expected out of promoting HMT technology. Moreover, the combination of MT and TM technologies is accepted to translate large amounts of standard documentation. Each of these technologies solves different subtasks within the general task: TM databases fulfill retrieval and substitution of previously translated content when MT provides with translation of a new one. Thus, such a convergence gains in high speed of translating any kind of content with minimal post-editing.

2. Three on-line translators under study differ both in features of their realization and in characteristics of produced output texts. The undertaken juxtaposition—corroborated by authors' practical experience—shows the following:

– no-one of these on-line translators is not able to translate the *whole* presented text correctly, the greatest challenge being the Russian case pattern;

– PROMT produces the majority of ludicrous translations as compared with other on-line translators, however being somewhat better at handling the *structure* of long compound sentences as the afore-cited sample has displayed;

– Google Translate shows better results in general rather than PROMT and provides a greater percentage of relatively correct translations of sentences;

– Yandex.Perevod approximates to Google Translate in translation quality both in rendering the sense and the structure of presented sentences, being rather efficient while processing large technical texts abundant in set expressions (its output texts need less post-editing as the authors' translate practice evidences).

Despite all developers' advertising claims we cannot take seriously the ability of on-line PROMT translator to compete against a human in practice yet. Translations produced by this system can give rather vague ideas of translated text content and they are often hard to post-edit. Perhaps the most promoted in Russia nowadays Google Translate is still far from being perfect giving rise to complaints of professional translators [6] but helps unsophisticated users to understand text fragments in an unknown language at large. Yandex.Perevod is a relatively new service but it makes progress in solving MT problems catching up with Google.

3. As on-line translators provide in general with fairly low grade output texts the latter obligatory need post-editing by a human. Furthermore, we should not forget that any documents implying legal liability such as contracts or warranties require verification by a specialist good both at languages and the subject matter.

Список литературы

1. Грабовский В.Н. Технология Translation Memory // Мосты. – 2004. – № 2. – С. 57-62.

2. Исламов Р.С., Фомин А.Г. Анализ современных систем машинного перевода типа SMT и RBMT // Филологические науки. Вопросы теории и практики. – 2013. – № 3 (21), Ч. I. – С. 69-73.

3. Машинный перевод [Электронный ресурс] // Компания ПРОМТ / © ООО «ПРОМТ», 2014. – Режим доступа: http://www.promt.ru/company/technology/machine_translation/. – (Дата обращения: 15.03.2014).

4. Машинный перевод [Электронный ресурс] // Компания Яндекс / © ООО «Яндекс», 2014. – Режим доступа: http://company.yandex.ru/technologies/translation/. – (Дата обращения: 10.03.2014).

5. Селиванов А. Настоящее и будущее технологий машинного перевода Google [Электронный ресурс] // Newsland / © Newsland, 2014. – Режим доступа: http://newsland.com/news/detail/id/599364/. – (Дата обращения: 12.03.2014).

6. Технология Translation Memory [Электронный ресурс] // Бюро переводов «Альба Лонга» / © «Альба Лонга», 2014. – Режим доступа: http://alba-longa.ru/translationmemory/. – (Дата обращения: 10.03.2014).

7. Brown P.F., Della Pietra V.J., Della Pietra S.A., Mercer R.L. The mathematics of statistical machine translation: parameter estimation // Computational Linguistics. – 1993. – Vol. 19, Issue 2. – P. 263-311.

8. Hutchins W.J. Machine Translation: Past, Present, Future. – Chichester (UK): Ellis Horwood, 1986. – 382 p.

9. Priyanga S.R., Azhagu Sindhu A. Rule Based Statistical Hybrid Machine Translation // International Journal of Science and Modern Engineering. – 2013. – Vol. 1, Issue 5. – P. 7-10.

10. Somers H.S. Example-based Machine Translation // Machine Translation. – 1999. – No. 14. – P. 113-157.