JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS

VOLUME 04, ISSUE 11
MOPNTHLY JOURNALS

ISSN NUMBER: 2751-4390
IMPACT FACTOR: 9,08

ALGORITHM OF MULTILINGUAL ELECTRONIC TRANSLATION

Uzakova Mamura Abduraimovna
Asia international university

Abstarct: Multilingual electronic translation is a crucial component of modern applied linguistics and artificial intelligence, enabling effective communication across languages in a globalized digital environment. This paper examines the algorithm of multilingual electronic translation as a multi-stage process that integrates linguistic analysis and machine learning techniques. The proposed algorithm includes input data analysis, tokenization and segmentation, linguistic processing, selection of an appropriate translation model, translation generation, post-processing, and quality evaluation. Special attention is given to neural machine translation models based on the Transformer architecture, which are widely used in contemporary multilingual systems due to their ability to preserve contextual and semantic relationships across languages. The study highlights the importance of post-processing and evaluation metrics in improving translation accuracy and fluency. The presented algorithm demonstrates the effectiveness of modern neural approaches in supporting multiple languages while maintaining translation quality and stylistic consistency.

Key words: multilingual electronic translation, applied linguistics, a multi-stage process,

Introduction

Multilingual electronic translation is one of the key tasks of modern applied linguistics and artificial intelligence. It is used in machine translation systems, voice assistants, search engines, and international communication platforms. The main goal of such algorithms is the automatic conversion of text from one language to another while preserving meaning, structure, and stylistic features[1-3]. At the first stage, the system receives the source text and performs its preliminary analysis.

Main tasks of this stage include:

1. identification of the input text language;
2. text cleaning (removal of unnecessary characters, HTML tags, and emojis);
3. normalization (bringing words to their standard form).

Tokenization and Segmentation. At this stage, the text is divided into smaller units [4].

Performed actions include:

1. splitting the text into sentences;
2. dividing sentences into tokens (words, subwords, symbols);
3. taking into account the specific features of a particular language (hieroglyphs, agglutination, morphology).

Sentence segmentation involves identifying sentence boundaries within the text. The system determines where one sentence ends and another begins, typically using punctuation marks such as periods, question marks, and exclamation points [5-7]. However, this process is not trivial, as punctuation can also appear in abbreviations, numbers, or quotations. Therefore, language-specific rules and statistical or neural models are often applied to correctly detect sentence boundaries. Proper sentence segmentation ensures that the translation model processes complete semantic units, which improves translation coherence and contextual accuracy.

Tokenization is the process of breaking each sentence into smaller elements called tokens. These tokens may represent full words, subword units, or individual symbols, depending on the language and the chosen tokenization method. Modern neural translation systems often use

JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS
VOLUME 04, ISSUE 11
MOPNTHLY JOURNALS
ISSN NUMBER: 2751-4390
IMPACT FACTOR: 9,08

subword tokenization techniques, such as Byte Pair Encoding (BPE) or WordPiece, to handle rare words, compound forms, and morphological variations. This approach reduces vocabulary size and allows the model to generalize better across different word forms. Accurate tokenization is crucial, as incorrect token boundaries can lead to loss of meaning or grammatical errors in the translation[8].

Different languages have unique structural characteristics that significantly influence tokenization and segmentation. For example, languages using hieroglyphic or logographic writing systems, such as Chinese or Japanese, do not employ spaces between words, requiring specialized segmentation algorithms. Agglutinative languages, such as Turkish or Finnish, form words by combining multiple morphemes, which may represent grammatical functions or meanings. In such cases, subword or morpheme-based tokenization is particularly important. Additionally, languages with rich morphology require careful handling of inflections and word forms to preserve grammatical and semantic information. Accounting for these language-specific features enables the translation system to produce more accurate and natural translations[9-11].

This stage is aimed at understanding the grammatical structure of the text. It includes:

1. morphological analysis (parts of speech, word forms);
2. syntactic analysis (relationships between words);
3. semantic analysis (meanings of words and phrases in context).

Modern systems use various approaches:

1. Rule-based translation — based on dictionaries and grammatical rules;
2. Statistical translation — uses probabilistic models;
3. Neural machine translation (NMT) — based on neural networks and the

For multilingual systems, a unified neural model capable of working with dozens of languages is most often used.

At this stage, the actual translation is performed.

The process includes:

— matching semantic units between languages;
— generating text in the target language;
— considering context, idioms, and set expressions.

Neural models use attention mechanisms, which allow them to capture relationships between words over long distances.

The algorithm for implementing electronic translation is presented below:

1-step.Beginning of the program;
2-step.Input text the source text in any supported language;
3-step.Language identification automatic detection of the language of the text;
4-step.Pre-processing cleaning and normalization of the text into a standard format;
5-step.Tokenization and segmentation dividing the text into logical elements;
6-step.Linguistic analysis analysis of the grammatical structure and semantic meaning of the text;
7-step.Translation model selection determining the optimal translation algorithm;
8-step.Translation generation producing text in the target language;
9-step.Post-processing improving the readability and correctness of the translation;
10-step.Quality evaluation assessing the accuracy of the translation;
11-step.Output translation the final result provided to the user;
12-step.Finishing of the programm.

After generating the translation, the system improves the quality of the output text.

This stage includes:

— correction of grammatical errors;

— restoration of punctuation;
— stylistic refinement;
— text formatting.

The final stage is the assessment of the quality of the result.

Evaluation methods include:
- ✓ automatic metrics (BLEU, METEOR, TER);
- ✓ comparison with reference translations;
- ✓ user evaluation.

Conclusion

The algorithm of multilingual electronic translation is a complex multi-stage process that combines methods of linguistics and machine learning. Modern neural approaches make it possible to achieve high translation quality and support a large number of languages, making such systems indispensable in the context of globalization and digital communication.

References

1. Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L.
The mathematics of statistical machine translation: Parameter estimation.
Computational Linguistics, 19(2), P- 263–311. 1993
2. Koehn, P. Statistical Machine Translation.Cambridge: Cambridge University Press. 2010
3. Vaswani, A., Shazeer, N., Parmar, N., et al. Attention Is All You Need.
Advances in Neural Information Processing Systems (NeurIPS), P - 5998–6008. 2017
4. Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate.Proceedings of ICLR. 2015
5. Sutskever, I., Vinyals, O., & Le, Q. V. Sequence to sequence learning with neural networks. Advances in Neural Information Processing Systems, P -3104–3112. 2014
6. Koehn, P., & Knowles, R. Six challenges for neural machine translation.
Proceedings of the First Workshop on Neural Machine Translation. 2017
7. Wu, Y., Schuster, M., Chen, Z., et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.
arXiv preprint arXiv:1609.08144. 2016
8. Jurafsky, D., & Martin, J. H. Speech and Language Processing (3rd ed.).
Stanford University. 2023