



PROMT Translation Solution for TripAdvisor

About TripAdvisor

TripAdvisor® is the world's largest travel site, enabling travelers to plan and have the perfect trip. TripAdvisor offers trusted advice from real travelers and a wide variety of travel choices and planning features with seamless links to booking tools. TripAdvisor-branded sites make up the largest travel community in the world, with more than 60 million unique monthly visitors, and over 75 million reviews and opinions.

Project Objective

Millions of user reviews comprise the core content of the TripAdvisor website. Many users worldwide would like to read reviews in their native language so the user-generated content needs to be translated for the localized websites. Traditional human translation is impractical to implement due to the huge volume of TripAdvisor content. To address this challenge, TripAdvisor has turned to machine translation (MT).

Due to the growing potential of the Russian travel market, TripAdvisor created a Russian version of its website and decided to translate all English reviews into Russian. PROMT, a leading provider of industrial MT solutions, was selected as the MT vendor for Russian.

Initial Requirements

The translation solution for TripAdvisor needs to meet three requirements:

- 1. High translation quality sufficient for understanding without human post-editing**

Machine translation of user reviews should produce clear and understandable content. Due to the huge amount of reviews, human post-editing of every single review is impractical.

It is important to note some natural challenges for any MT system with respect to TripAdvisor's content:

- User-generated content is a challenge for MT since such texts are typically informal and often contain spelling, grammar, and punctuation errors that affect machine translation at the crucial first step of source text processing.
- Another challenge is that at the start of the project, TripAdvisor did not have a large and relevant parallel corpora for EN>RU engine training, so there were not enough data for a pure statistical approach.

- 2. Automatic evaluation of translation quality**

Evaluation of the actual quality of output is a very important aspect for any MT solution. Since manual evaluation of the entire translated TripAdvisor content is impossible, the MT solution must provide an automatic scoring mechanism to evaluate the quality of the translated texts. PROMT needed to develop a confidence score of the MT output to enable TripAdvisor to publish only high-score translated reviews.

- 3. Integration into the workflow of TripAdvisor's website**

TripAdvisor website developers wanted to use MT as a web-service on a remote server that receives translation requests and returns responses in a format which requires minimal post-processing on TripAdvisor's side.

Solution Deployment

In May 2011, PROMT translated a batch of high-priority reviews with a baseline out-of-the-box MT engine. Once the baseline quality was evaluated, the solution deployment started with an in-depth customization of the PROMT baseline engine to specifically translate TripAdvisor content.

In November 2011, the PROMT translation solution was fully integrated into TripAdvisor's workflow and since then, around user reviews are translated on a weekly basis.

The MT server is hosted in the PROMT datacenter and is fully maintained by PROMT.

TripAdvisor continues to provide PROMT with additional linguistic data for further engine training and the translation quality continues to improve.

Components of the PROMT solution for TripAdvisor

The main components of the PROMT solution customized for TripAdvisor:

- [PROMT Translation Server 9.5 DE](#) – a reliable, robust, and scalable server-based solution that allows the translation of large text volumes
- [PROMT DeepHybrid](#) – a technology that produces high-quality MT that is comprehensible for end users

The following components were included in the PROMT DeepHybrid engine for TripAdvisor:

- **Specialized Dictionaries:** For this project, PROMT revised its "Travel" dictionary for TripAdvisor. In addition, **an advanced pre-processing of source errors** was developed. The translation system is able to recognize errors, misprints, and abbreviations to preserve the meaning of the original text.
- **Language Model:** PROMT collected and processed a special Russian travel-related user-generated corpus. That corpus was used to build a **language model** which is necessary for the statistical post-editing of the translated reviews and for the automatic evaluation of their translation quality.
- **Statistical Post-editing:** This feature allowed PROMT to improve the initially trained engine on a regular basis. The reviews which are most frequently read by Russian users go through human post-editing and become the basis for a translation correction database (**statistical post-editing table**) which makes the translation smoother and more human-like.
- **Automatic evaluation of translation quality** (confidence score). Right after a review is translated, the PROMT engine automatically compares the resulting text against the **language model**, producing a metric of how close the Russian translation is to a native Russian corpus.

To define the confidence score threshold suitable for TripAdvisor's content, PROMT compared the automatic confidence scores to a human evaluation of translations performed by multiple linguists over a sizeable random sample of reviews.

- **Dedicated web-service for TripAdvisor.** The solution was deployed on a dedicated server in the PROMT datacenter. At the integration stage, PROMT and TripAdvisor agreed on the most convenient way for the TripAdvisor developers to format the data exchange and PROMT implemented a dedicated API function and made it available for TripAdvisor.

Conclusion

The PROMT Translation solution for TripAdvisor fully addressed the project objectives and initial requirements and has become a solution that has helped TripAdvisor to translate a large amount of reviews into Russian.

Implementation Results:

- Large text volumes are translated quickly.
- The translation quality is sufficient for full comprehension.
- The translation costs of the entire content are considerably less than the human translation costs of even a small fraction of reviews.
- An automatic scoring of translation quality has been implemented.
- The solution has been integrated into the TripAdvisor workflow with minimal support overhead on TripAdvisor's side.

"The TripAdvisor website is very popular due to users who write about everything interesting to travelers: hotels, restaurants, places of interest, etc. Therefore, it is a business-critical task to make user reviews available in different languages for travelers all over the world. Because of the large amount of reviews, we can use only high-quality MT to address this challenge. We have selected PROMT to translate from English to Russian and this has been successfully working for the last year. PROMT helps us to deliver the experiences of English-speaking travelers to our Russian-speaking audience."

- Lorna Whelan, Senior Manager, Localisation, TripAdvisor

"Such a large website as TripAdvisor, which aggregates a huge amount of user-generated content, is a perfect customer for MT and PROMT DeepHybrid technology. A statistical approach works well only if there are large volumes of parallel data for training, while the PROMT DeepHybrid approach produces high-quality translation even when the training data are very limited. Customizing our MT engine for TripAdvisor was a demanding and interesting challenge because of the large volume and the specific nature of the content. I think we did it well, and now, millions of Russian-speaking Internet users can plan their trips based on the experiences of other TripAdvisor users worldwide."

- Julia Epiphantseva, Business Development Director, PROMT