

Applying Machine Translation to an Online Travel Localization Strategy Effectively

TABLE OF CONTENTS

Introduction.....	1
Time to Publish in an Online Economy.....	1
The Emergence and Importance of User Generated Content.....	2
Online Travel Machine Translation Use Cases.....	2
A Machine Based Localization Solution for the Travel and Tourism Industry.....	4
About Omniscien Technologies.....	6

INTRODUCTION

The travel industry is one of the global industries in which multilingualism is a fact of life. With more and more travelers taking travel matters into their own hands, online channels increasingly gained unopposed significance. Whether travel suppliers want to provide contents of various interesting travel destinations for their outbound travelers or want to reach out to potential inbound travelers in different countries to invite them for a special offer, or for booking a hotel room or trip, visiting a specific attraction or event, etc. the core for all this is communication and related to communication – is localization.

TIME TO PUBLISH IN AN ONLINE ECONOMY

Online as a channel has become a vital tool for travel service providers. It now serves as one of the main advertising spaces as well as a touch point for sales and customer service. Managing all the online content, including all its constant updates, and making them accessible in the target market’s language in near-real-time is imperative. The conventional approach of localising would be a human only translation, which is often time consuming and costly. However, in a fast-paced business as the travel sector, time is an important factor and localizing content in a timely manner in order to make them INSTANTLY available is essential in order to stay competitive. This requires segmenting content types and applying human translation wherever high quality is required, time is less of a factor and if it is economically feasible. However, irrespective of the localization need, drawing on technology effectively to support localization is of strategic importance. While machine based



HIGHLIGHTS

- As time to publish and cost become ever increasing factors in an online economy, effective content segmentation and technology application are key success factors.
- To have contents such as hotel/sites reviews, user comments and other user-generated content published across languages fast in order to provide it in the target user's language is crucial.

solutions can assist the human localization process, for some use cases machine translation technology is the only viable option. For example, live chat solutions for providing multilingual customer and sales support, be they bot or human supported, to support global customers can only be enabled by real-time machine translation.

As time to publish and cost become ever increasing factors in an online economy, effective content segmentation and technology application are key success factors.

THE EMERGENCE AND IMPORTANCE OF USER GENERATED CONTENT (UGC)

User-generated content (UGC) as peer provided information for Travelers has become a must for online services. Hotel/Sites reviews as well as user comments are among favourite reading before booking a room or planning a trip, and unsurprisingly influence travelers' buying decisions. To have such content published across languages fast in order to provide it in the user's language is crucial. However, localizing UGC is challenging due to the high volumes (large providers will have billions of words a day across multiple languages), due to the fact that it needs to be localized cost-effectively in near real-time, and last but not least due to the fact that the quality of the source content is often of different quality. In such a case, a highly sophisticated machine translation solution can be the answer.

ONLINE TRAVEL MACHINE TRANSLATION USE CASES

The range of use cases is broad and increasing, but the following list provides some of the more common use cases:

- Localization of Traveler Reviews

A large number of travel sites rely on social media integration and traveler reviews. There are also many sites that take reviews from high volume markets such as the US and localise them to smaller markets in order to bolster their website and increase attractiveness to local customers.

HIGHLIGHTS

- Only few solutions are capable of handling textual data when it comes to big data analytics and capabilities of supporting different languages and entity processing are required for this task. In order to provide desired insights, a combination of language processing, machine translation and machine learning is necessary.

- Localization of Special Offering and “Low Value” content

Online Travel sites usually distinguish between high value-offerings attached to them such as descriptions of luxury hotels and top destinations. Often, complicated analytics are behind the scenes in order to make sure that traffic is “sticky” and revenue is maximised. However, the same Online Travel sites mostly have “lower value” content as well, which might drive less revenue, but nevertheless are significant as generator of revenue. Falling into this category could be short term special offers that have a short life span on the web. For this kind of content costly human translation might not be required as it could be done faster and more economically with machine translation.

- Customer Support and ChatBot Localization

To support a multi-lingual clientele a customer support human agent needs to be able to understand and communicate in many different languages. Especially ChatBot solutions require real-time, high quality translation in order to do the service. This is a classic use case for professional machine translation to support customer service.

- Big Data Analytics

According to EyeforTravel’s “The State of Data in Travel” Report 2017, most travel organisations (65%), now have a dedicated data, analysis or insight team, who analyse data through diagnostic, descriptive, and predictive analytics. However, the report also states that there “remains work to do as nearly two thirds (64.4%) are not yet able to use their data to build prescriptive analytics, the most sophisticated form of analytics, which can give travel business a complete outlook and strategy”. (from “The State of Data in Travel”, EyeforTravel Report 2017).

Only few solutions are capable of handling textual data when it comes to big data analytics. Furthermore, capabilities of supporting different languages and entity processing are required for this task. In summary, in order to provide these desired insights, a combination of language processing, machine translation and machine learning is necessary.

A MACHINE BASED LOCALIZATION SOLUTION FOR THE TRAVEL AND TOURISM INDUSTRY

As for any business process support systems and IT solutions, a thorough understanding of the use case and requirements are critical in order to address the needs and requirements of the travel industry. This understanding must embrace all dimensions ranging from data processing workflow, system integration requirements, specifications and knowledge of the type and format of the content to be processed, as well as the required output quality and format and financial parameters related to the project. This means it is a localization as much as an IT challenge and requires the know-how to handle this accordingly.

The following steps form a high-level guide on how to proceed with such projects.

1. Understand the use case and its dimensions and build the right team

Every use case is unique and while similar building blocks can be re-used, understanding the use case end to end is key to success. Understanding system integration, workflow, localization and conversion aspects, content sources, quality requirements, volumes, and cost parameters are just some of the details required and will also help understand the skills the project requires to succeed. When bringing in external partners, define their roles and integration points clearly and ensure they have the required subject matter expertise to perform their task and interact with your team effectively.

2. Design the architecture, i.e. workflow and the machine translation engine and system requirements

Don't cut corners in the early stages of the projects; having well defined requirements and a solid architecture that is fit for purpose is a fundamental pre-requirement for success. As part of the architecture and design also include all the analysis related to data; for example, what meta data, source and target data will be provided and/or is needed to adequately process the content. Also, consider what style guides and glossaries exist and need to be used? Do entities need to be recognized and processed in a specific manner or does the system need to score the content quality?

3. Develop a custom engine and workflow and ensure the required mode and capacity are available

Once the requirements are understood and the architecture and data sources have been defined, the custom workflow

and engines can be implemented. In order to achieve quality output, professional customization is required. This means that the workflow, the non-translation capabilities and the engines themselves need to be customized.

4. Test

Test, improve and test again. With complex content, such as travel related text and specifically UGC, testing is required to ensure that all data formats are processed correctly before going into production.

5. Monitor and create a feedback loop

Once released, the work is not done. Source content changes over time, terms change, locations change, patterns change and continuous QA is required to monitor quality. Ensure that there is a proper feedback loop. Often translation quality will improve fast if possible issues from an initial engine and workflow release are found during the process of volume data production and addressed correctly as well as if engines are re-trained. While efforts are largely “front loaded”, over time the required efforts will be reduced as the system matures.

Hence, for localising large content volume where time to market and cost effectiveness are pivotal in order to be in the vanguard, drawing on machine translation and language processing technology could serve the purpose.

ABOUT OMNISCIENT TECHNOLOGIES

Omniscien Technologies is a leading global supplier of high performance and secure high-quality Language Processing, Machine Translation (MT) and Machine Learning technologies and services for content intensive applications. Our wide range of solutions serves clientele from various industries including the Localization Industry, Online Research Services, Publishing, eCommerce, Media, Online Travel, Technology, Enterprise and Government.

Omniscien Technologies has gained a reputation for cutting edge solutions with its Language Studio™ platform. Depending upon the customer's unique requirements, Language Studio™ can be deployed in a variety of ways to integrate with their in-house data processing and translation management systems, and it offers unparalleled levels of customization and control as well as feature rich pre- and post-processing, enabling customers with even the most complex data to achieve both high quality and high-volume output to satisfy every use case. Omniscien Technologies has by far the most comprehensive and feature rich system in the market today.

Covering 550 language pairs and with a number of industry specific solutions, Omniscien Technologies remains the partner of choice for customers with complex, high-volume bespoke data processing and machine translation needs.

For further information on Omniscien Technologies or Language Studio™, please visit www.omniscien.com or contact sales@omniscien.com