Machine Translation - Quantum Leap or Flash in the Pan

Machine Translation (MT) simply put is the use of software to translate text or speech from one natural language to another. MT performs simple substitution of words in one natural language for words in another, but that alone usually cannot produce a good translation of a text because recognition of whole phrases and their closest counterparts in the target language is needed. To process any translation, human or automated, the meaning of a text in the original (source) language must be fully restored in the target language, i.e. the translation. While on the surface this seems straightforward, it is far more complex. Solving this problem with corpus and statistical techniques is a rapidly growing field that is leading to better translations, handling differences in linguistic typology, translation of idioms, and the isolation of anomalies. MT software is customized by domain or profession (such as weather reports), improving output by limiting the scope of allowable substitutions. This technique is particularly effective in domains where formal or formulaic language is used. It follows that machine translation of government and legal documents more readily produces usable output than conversation or less standardized text. Machine translation with no human involvement was pioneered in the 1950s and has come a long way in the last 60 years.

Primarily, there are two types of automated or instant translation software, rule based and statistical. Rule-based systems use a combination of language and grammar rules plus dictionaries for common words. Specialist dictionaries are created to focus on certain industries or disciplines. Statistical systems have no knowledge of language rules. Instead they “learn” to translate by analyzing large amounts of data for each language pair. They can be trained for specific industries or disciplines using additional data relevant to the sector needed.

Machine Translation has made considerable advances in terms of accuracy, consistency and even fluency and flow of language when translating from one natural language to another. However when, what type of, how, where MT is to be used needs to be carefully considered before using MT given the various aspects and issues involved in it.

Machine translation does have significant advantages like speed, ability to process large volumes and low cost of translation, there are some tough challenges in using MT given the fact that no two languages are alike in terms of structure, grammar, use of words and many other aspects.

Given this background the current Case in Information Systems is being presented. The facts of the case are based on information available in media reports, online information and some real life incidents. Although every case may cover multiple aspects it will have a predominant focus on some aspect which it aims to highlight.

A case study cannot and does not have one right answer. In fact answer given with enough understanding and application of mind can seldom be wrong. The case gives a situation, often a problem and seeks responses from the reader. The approach is to study the case, develop the situation, fill in the facts and suggest a solution. Depending on the approach and perspective the solutions will differ but they all lead to a likely feasible solution. Ideally, a case study solution is left to the imagination of the reader, as the possibilities are immense. Readers’ inputs and solutions on the case are invited and may be shared. A possible solution from the author’s personal viewpoint is also presented.

A Case Study of Kachwala Mistry & Partners

Sameer Kachwala is the Senior Partner of Kachwala Mistry & Partners. Of late the firm is facing a lot of difficulties in filing plaints, registration papers and other legal work on time. One of the causes of concern is the inordinate delay in getting the documents translated. The State of Anarthapur and its neighboring State Nirdhanabad, where the firm’s practice is primarily concentrated, both have nine major languages of which two are used in the courts and are official state languages for documents, land records and legislation. The High Court of both states also uses English as its official language. The firm being an international law firm with clients mainly in UK and US uses English to transact its business.

The problem of delay had arisen due to a series of events – first a strike in the official translation department of Nirdhanabad, followed by a language agitation in Anarthapur, which was further compounded by the fact that two of the senior translators who were with the firm for over 30 years had retired and their services were no longer available. The firm had tried different alternatives including outsourcing work to external agencies but somehow things were not working out.

Sameer and his partner Adi had just returned from a business trip overseas, where they had come across Machine Translation software and were very impressed. Determined to do something about the situation both Sameer and Adi have decided to introduce MT and call a meeting of all partners and senior staff including the IT manager.
Sameer is quick to list the advantages:
- MT is fast - When time is a crucial factor, machine translation can really help. You don’t have to spend hours poring over dictionaries to translate. Instead, the software provides quality output in no time.
- MT is economical - It is comparatively cheap. There is an initial investment but in the long run it is a very small cost considering the return it provides. If a professional translator is used, he will charge you on a per page basis which is extremely costly while software has one time cost.
- MT can deal with Multiple Languages - In a state using nine languages the same software can simultaneously provide translation from / to any / all the languages – whereas for each combination a different professional translator may be required.
- MT can process large volume of data - A heavy work load will result in piling up work, backlog and delays in manual translation but not so when MT software is used.

Adi added that there could be some issues about quality, accuracy of translation or some such issues but was sure that the experienced staff will be able to deal with it. “After all we do review the translations before submitting them.” – He had remarked.

Predictably Anil, Makarand and Hemant all in their thirties had jumped at the idea. So was the IT manager elated. The resistant lot seemed to be Dwarkanath who belonged to the old guard and Yogendra the Staff representative.

Yogendra was outspoken and raised the issue of the three vacancies in the translation department which had not yet been filled and attributed the problems to this. He was also apprehensive that with the MT software, not only these will not be filled but the five staff currently employed in the translation department would also be asked to go.

Dwarkanath had more fundamental issues. He stressed the ambiguity and gaps in words in different languages. The word run in English has more than fifty different meanings and usages and so were many other words in all languages. This fangled software does not understand the context and the fine nuances of language.

He gave an example of difference in syntax in languages with the following example of the result he produced after using the MT software to first translate from English into French and then back, which produced results which are completely different –
- She ran into the room. (English)
- Elle entra dans la salle en courant. (French)
- She entered into the room in/while running. (English)
- Ea intra in camera alegând. (French)

Google Translate helps paramedics deliver Baby - February 10, 2015

The Swahili word for “thanks” might be appropriate to Google Translate, after it helped two paramedics deliver a Congolese woman’s baby in Ireland this week. Men’s lifestyle website Joe.ie cited a report by The Corkman news site that the incident occurred this week between Macroom and Lissarda, when the paramedics were bringing the woman to a hospital to deliver her child.

“It’s something that I think I won’t ever forget as I was translating Swahili into English somewhere on the side of the road between Macroom and Lissarda,” paramedic Gerry McCann said. McCann and Shane Mulcahy, were taking the woman to the Cork University Maternity Hospital when the baby was due. The problem, however, was that the woman spoke limited English. Thinking quickly, McCann opened Google Translate on his phone to communicate with the woman. As a result, a “beautiful baby girl” was born, the report said.

Sameer and Adi who had already made up their mind grabbed this opportunity and pushed through the change of adopting MT software unanimously at the meeting. However realizing that there was indeed merit in the issues raised in the meeting they thought it wise to consult Radha who had extensive experience in MT software and its implementation, to guide them in this exercise.

Radha has a series of meetings both group and one on one with the partners, different departments and staff and has now come up with an understanding of the major issues. What would be your thoughts if you were Radha?

Solution

The situation
The firm is currently facing several issues in meeting deadlines, filing plaints, affidavits and preparing legal documentation. One of the main reasons is the backlog in translations. The firm operates in an environment using multiple (up to nine) local languages as well as English. The backlog has been aggravated due to strike in the Translation department of the court and other issues which have put pressure on the translation staff of the firm. This department is currently handicapped due to the retirement of two senior translators for whom no replacement has been found.

The situation is likely to increase in the long run with increasing costs of translation, difficulty in outsourcing translation of confidential and sensitive documents and ever shortening deadlines and increased expectations of speed and quality of output - given the current level of competition and globalization.

Machine Translation provides distinct advantages like speed of processing, ability to translate large volume of text / matter, comparatively cheaper cost per page of translation, ability to simultaneously deal with multiple languages and above all improved translation ability with the recent advances in MT software. Adoption of MT software seems to be the option of choice as it does provide a viable alternative to supplement and strengthen the present process of manual translations.

However, as noted there are many limitations and short comings of MT software. There are quality issues, issues of inability to translate effectively where structural or linguistic differences exist and where context provides the meaning. The level of accuracy expected in a law firm is of professional standard and a MT software translation may not be able to meet or even come near to that quality.

The consequences
The issues and challenges and even the resistance within the organization is being aggravated not by the proposed adoption of MT software but the manner in which it is proposed to be used and done. Unless the issues and challenges are met and overcome and the resistance within is addressed satisfactorily, the firm will end up with more problems than before post implementation of MT software.

The Strategy
The right strategy for Kachwala Mistry & Partners, the law firm, at this stage would be: Identify and address the issues challenges and resistance systematically:
Issues and Challenges -
1. MT software output is not consistent in quality
2. MT software overlooks / cannot understand context which has a significant impact on meaning
3. MT software is unable to effectively deal with differences in language structure, differences in construction, words having multiple meaning and usage, idioms and phrases, structural bilingual ambiguity, lexical differences and a whole lot of linguistic issues.
4. Translation is not merely word replacement and MT software cannot and does not take into account the customary usages and body of knowledge and conventions specific to particular languages.

Resistance from within -
1. Perceived curtailment of jobs and redundancy in the once strong translation department
2. Inability of software to address language issues
3. Possibly multiple languages for which Translation software is not developed / available and the fear of being unable to cope with the new system especially of the older employees.

Adoption of Hybrid Approach: The right approach to be adopted in this situation will be to adopt a human / machine compromise - a hybrid approach. Machine translation with what we call pre- and post-editing is a methodology in which a linguist “trains” or programs the machine-translation engine to correctly translate context-specific terminology, phrases with double meanings and case-based client exceptions to rules where the MT platform may have otherwise made a mistake. The content is then processed by the machine translation software and then after translation, a professional human translator reviews the output and edits it for technical accuracy, style and comprehensibility.

Providing MT software to supplement and strengthen the Translation Department and not as a replacement: Providing MT software as a tool will remove staff resistance bred on fears of replacement / redundancy. It will also address the concerns of quality of translation as it will be reviewed. Introducing the human element thus removes most of the major issues, challenges and resistance.

Creation of Decision Rule for assigning jobs to MT software: Decision rules based on criteria covering different aspects like volume of work, nature of translation document, intended use internal / external - certain jobs can be entirely assigned to MT software with limited review. Others which do not meet criteria or are difficult will see limited use of MT software with substantial human intervention.

The way forward is to adopt the hybrid approach to improve the working of the Translation Department, with MT software being acquired and introduced not to replace it but to supplement and strengthen its working and improving the overall efficiency and effectiveness of the firm. This way, the firm can balance the need for speed and cost benefits of machine translation and address the potential pitfalls.

An effective solution is generally expected to proceed on these lines.

About the Author
Dr. Vishnu Kanhere is an expert in taxation, fraud examination, information systems security and system audit and has done his Ph.D. in Software Valuation. He is a practicing Chartered Accountant, a qualified Cost Accountant and a Certified Fraud Examiner. He has over 30 years of experience in consulting, assurance and taxation for listed companies, leading players from industry and authorities, multinational and private organizations. A renowned faculty at several management institutes, government academies and corporate training programs, he has been a key speaker at national and international conferences and seminars on a wide range of topics and has several books and publications to his credit. He has also contributed to the National Standards Development on Software Systems as a member of the Sectional Committee LITD17 on Information Security and Biometrics of the Bureau of Indian Standards, GOI. He is former Chairman of CSI, Mumbai Chapter and has been a member of Balanced Score Card focus group and CGEIT- QAT of ISACA, USA. He is currently Convener of SIG on Humane Computing of CSI and Topic Leader – Cyber Crime of ISACA(USA). He can be contacted at email id vkanhere@gmail.com