Translator-Expert Retrieval in Multi-Aspect Framework

Mihai Lupu
Information and Software Engineering Group
Vienna University of Technology
Favoritenstrasse 9-11/188
A-1040 Vienna, Austria
Email: lupu@ifs.tuwien.ac.at

Navid Rekabsaz
Faculty of Informatics
Vienna University of Technology
Email: rekabsaz.n@gmail.com

Abstract*—TODO: The abstract goes here.*****

I. Introduction

*** TODO: Introduction to expert retrieval and learn to rank goes here ***

The designed framework can be categorized in domain of on-line translation. The framework aims to find the best translator-corrector combination for a client who seeks for a reasonable offer to translate a document.

The profile of every translator consists of different aspects like offered price per word, translation time per word, experience and quality of translation in a specific language pair (source and destination languages). Based on client's document, the system figures out the value of each aspect and offers the first three most related translators. As it is clear, the aspects are in different scales (price in currency, time in minute and etc.). It should also mentioned that the one with best offer in all aspects (lowest price, fastest time, most similar to previous experience and etc.) may not be the best offer and leads to a low quality translation and unsatisfied client.

In order to evaluate the quality of translators and also whole the task, a feedback system for corrector and client is considered. After finishing translator's task, the corrector starts revising the text and leaving a feedback on it. The feedback of corrector which is a number between one to five is used as a base for estimating the proficiency of translator. Then by delivering the task to the client, his feedback is asked. The feedback of client is used to deploy and leverage ranking system.

In Section II, the designed Translator-Expert Retrieval framework is discussed in detail. Then, Section III summarizes system evolution and achieved results. Finally conclusion is written down in section IV.

II. Translator-Expert Retrieval Framework

In this section, first the framework's elements is denoted. Then the techniques for calculating documents similarity and some restrictions are described. At the end, the strategy for ranking and choosing the best possible translator is discussed.

A. Framework Denotation

As it is depicted in the following formula ,the profile of each translator related to the query ($\overrightarrow{T_{i,q}}$) is a vector where i is the iterator for translators and q is the corresponding query. In the formula, t_q and c_q stand for estimated time and cost for translation. e_q denotes the experience or knowledge of translator which is calculated based on similarity between query document and all previous-translated documents. The technique applied for calculating e_q is explained further. s shows the quality of translator and how previous tasks have been estimated. It is achieved by calculating the average value of correctors' feedbacks. The value is between one (very bad) to five (perfect).

$$\overrightarrow{T_{i,q}} = \{t_{i,q}, c_{i,q}, e_{i,q}, s_i\}$$

After finishing the task and gathering feedbacks of clients, based on the offered values for each translation, feedback vectors can be built up. In the following formula j denotes the iteration for feedbacks and f_j stands for feedback value.

$$\overrightarrow{F_j} = \{t_j, c_j, e_j, s_j, f_j\}$$

B. Documents Similarity

In order to calculate e_q, we need to figure out the similarity between query document and all previous-translated documents. The similarity is calculated by Lucene library and is denoted as SIM function. Every pre-translated document is assessed by corrector's feedback (f'). In additional to these documents, every translator may submit some other documents which had been translated before and clearly not assessed. In the framework, we assume their quality as medium (two). In the following formula $d_iterator$ and d_q stand for pre-translated and query documents. n and m are the number of pre-translated document inside and outside the framework.

$$e_{i,q} = \sum_{0<k<n} f'_k SIM(d_k, d_q) + \sum_{0<l<m} 2SIM(d_l, d_q)$$

***TODO: The Lucene's similarity value does not have a special range and is a number greater than or equal to zero. Here there can be some discussions about the meaning of the achieved value or probable alternative solutions ***

C. Ranking

TODO: Ranking method goes here

III. Evaluation

TODO: Evaluation goes here.

IV. Conclusion

TODO: conclusion goes here.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.