

Health Translations

A crowdsourced, gamified approach to translate large vocabulary databases

Augusto Cravo Silva

Department of Informatics Engineering,
Faculty of Engineering, University of Porto
Porto, Portugal
augusto.cravo.silva@fe.up.pt

Carla Teixeira Lopes

INESC TEC and DEI,
Faculty of Engineering, University of Porto
Porto, Portugal
ctl@fe.up.pt

Abstract — The health domain is rich in specific vocabulary and information structures. Previous work on this area includes the collection of this information in information systems. However, the language of these can limit their use. To overcome this, we present Health Translations, a web application that uses crowdsourcing to translate a large vocabulary set that, currently, is only available in English. To increase usage, gamification methods are applied that reward both the *quantity* of collaboration and the *quality* of it. When completed, these translations can be made available without costs to the research community. This paper presents the platform as a responsive web application.

Keywords - health information retrieval, crowdsourcing, gamification, cross-language retrieval, interactive machine translation.

I. INTRODUCTION

Information processing can be improved through the use of existing *information structures*. Unfortunately, much of these structured data is only available in one language which leads to difficulties when applying the same techniques to other languages. The Health Translations web application proposes a solution to this problem using a crowdsourcing approach with gamification. Although it can be adapted for other information structures, it was originally developed having in mind the translation of the Consumer Health Vocabulary (CHV), a vocabulary linking everyday words about health to technical terms used by health care professionals [1], only available in English. The CHV includes more than 158,000 strings.

Crowdsourcing translation is a technique that allows the collaborative translation of information with professional quality by using ranking techniques and reducing the effort of each contributor [4] in what can be seen as a tedious task. To improve the motivation of the users, gamification is applied, i.e., “the use of game design elements in non-game contexts”[5]. Also, automatic translations are used as a starting point for the correct translation.

Besides its contribution to the translation of general information structures, Health Translations aims to make the CHV available in several languages and without costs to the research and medical communities. It is specially designed to be used by health care professionals.

This paper will provide an overview of the system and of the expected results.

II. RELATED WORK

A. Crowdsourced and Automated Translations

Crowdsourcing platforms are getting increasingly popular. One known example is Amazon’s Mechanical Turk [2], an online marketplace designed for people to complete *Human Intelligence Tasks* and get small fees by completing each task. More specifically, crowdsourcing translation platforms are also widely available these days. One example is Transifex [3] that make it easy to localize websites and applications both by professional methods and crowdsourcing ones. There are different roles such as translators and reviewers that ensure the quality of the contributions. The contribution of automatic translation methods to human translation services and vice-versa has been described by S. Green et al [7]. They developed a “mixed-initiative system in which human and machine agents interactively refine translations” [7] starting with an automatic translation that, when modified by the user, also suggests new translations for the remaining words, based on context.

B. Gamification

“Gamification has been defined as a process of enhancing services with (motivational) affordances in order to invoke gameful experiences and further behavioral outcomes” [8]. It can be broken down to three parts as seen in fig. 1. This way, when developing a gamified approach, one should:

- Choose which affordances to use. For example: points, leaderboards, achievements/badges, levels, story/theme, clear goals, feedback, rewards, progress;
- Understand the psychological outcomes of the gamification techniques in place: motivation, attitude and enjoyment;
- And, understand the behaviors resulting from these psychological effects.

C. CHV related work

CHV has been used for different projects related to Health Informatics. Josceli Tenrio et al. [9] also introduce a proposal for translating it. However, this work focus only in the translation to Brazilian Portuguese and relies on the Delphi method in contrast to crowdsourcing.

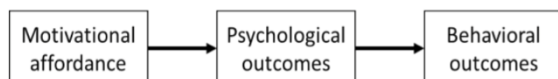


Figure 1. Main parts of gamification. [8]

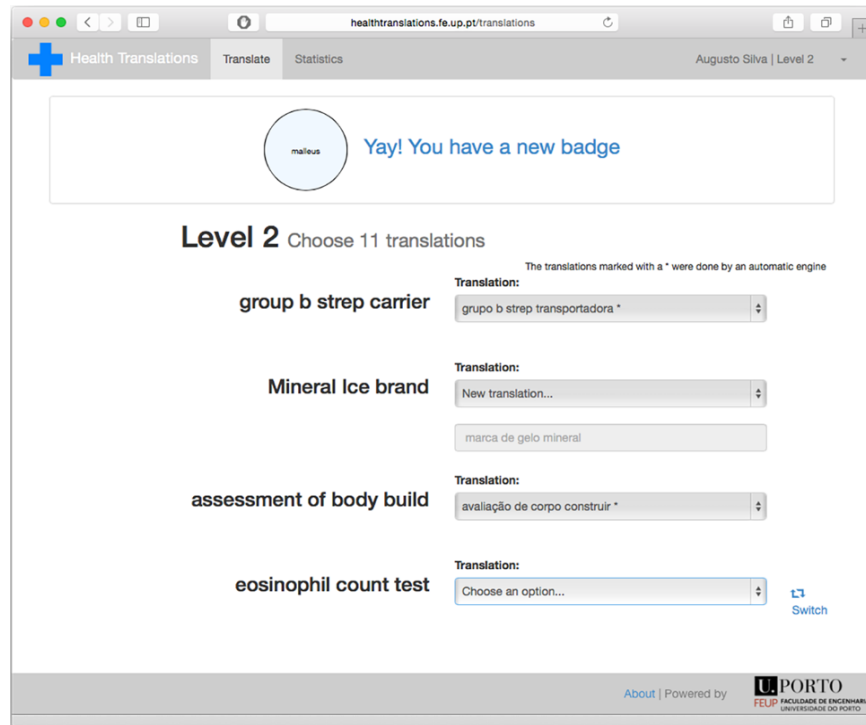


Figure 2. Voting page.

III. CROWDSOURCING TRANSLATIONS WEB APP

The system was developed in the form of a multi screen web application built with *Ruby on Rails* and with information persisted by *PostgreSQL*. Here the community is asked to translate medical strings, i.e. *vote*, either by selecting an existing one or by proposing new ones (that will be added to the list of available translations for that string). To speed up the process, we provide translations for the majority of the strings. These translations were previously obtained automatically through Google Translator and Wikipedia and are preloaded in the database alongside with the CHV strings.

When faced with a string to translate, users can ask that string to be switched to another. This is useful for situations in which users have doubts and avoids low-quality translations. Moreover, all new translations are spell checked to improve overall quality as the user types the words (see Section V). These features are illustrated in fig. 2.

In what interaction is concerned, the strings to be translated are presented in an incremental manner, i.e., only after the previous one is submitted but without leaving the same page; the automatic translations are clearly marked and unknown/incorrect words found by the spell checker appear below the input box. There are also interaction aspects related to gamification, which are exposed in Section III-A.

In Section III-B we explain the different ways the strings can be prioritized for translation and the criteria that makes a translation valid, so that, the corresponding CHV string is no longer showed to the users to be translated.

A. Gamification

To give users motivation to translate more strings, the system applies gamification techniques that not only reward the number of translations but also their quality. To remove the overhead of the user registering process, there is also the possibility to enter the web application as an anonymous user but this will prevent the user to take part in the gamified experience. The following subsections describe the used gamification techniques.

1) *Levels*: The level of a user is determined by the number of strings he/she has translated. As the level gets higher, the number of required translations increases. This feature aims to increase the number of translations per user.

2) *Points*: When a user votes on the same translation other users voted before, these other users get points. By doing this, the quality of their votes and their contribution to the validation of the translations is rewarded.

3) *Badges*: The number of provided votes and won points are then converted into badges as the user reaches predefined milestones. Since the community that will use this system is composed by health care professionals, badges will be associated with persons with outstanding achievements in the health domain (e.g. Nobel laureates). Since a part of the target users are still studying towards a medical degree, this will also have a pedagogic impact on them. An example of a badge being rewarded (to be improved in the future with a graphic visualization) can be seen in fig. 2.

4) *Ranking in the workplace*: Each user has a profile with his/her achievements and how these compare to their

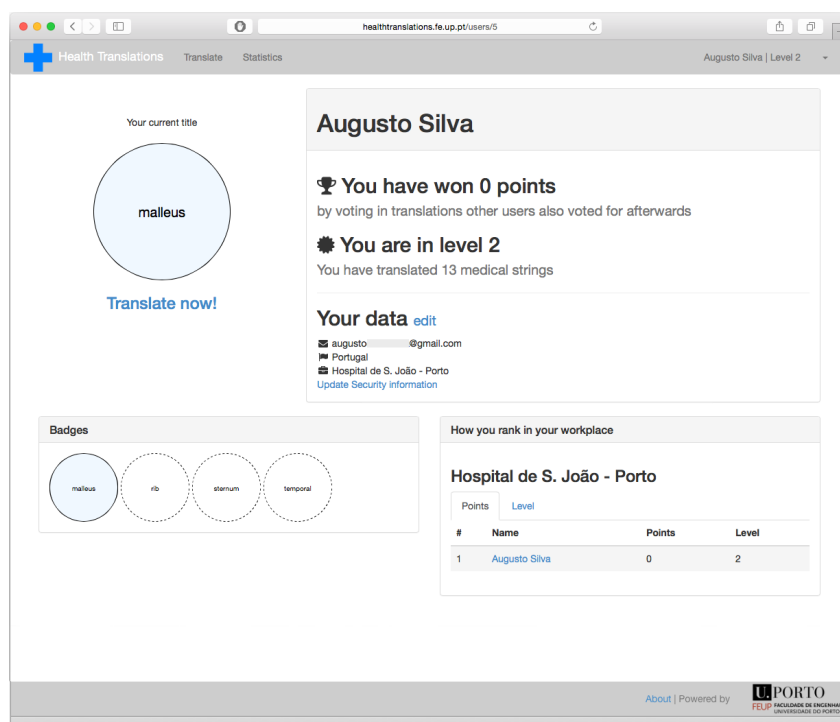


Figure 3. User Profile.

colleagues and by this take part in a friendly competition which is expected to increase usage. By placing the mouse over the badges users know what is the corresponding milestone. An example of a profile can be seen in fig. 3.

B. Statistics and Administration

From user interaction, there are some important statistics which provide the interested parties an overview of the platform: the progress of the project via the percentage of CHV strings that have been successfully translated; the accuracy of automatic translations by analyzing their number of points; the top strings users chose not to translate (i.e. clicked on the “switch” button when in the Voting page), which can be seen as difficult concepts. An example can be seen in fig. 4.

Considering these, the administrator can adjust how the application behaves by choosing:

- When a string is considered to be translated, i.e. by a minimum number of votes in the most voted option, a minimum difference of votes between the first and second most voted options or a combination of both;
- What is the current priority, i.e. to get the valid translation of some strings or to get some translations for all, which will change which strings are going to be translated next - the ones where the best translation has the highest number of votes or randomly, respectively;
- To hide non-relevant strings, e.g. chemical compounds, and to edit/hide incorrect/irrelevant translations.

An example of translation management can be seen in fig.

5.

IV. CURRENT WORK

Currently, we are testing the platform with some users and improving its features with their feedback. We are also improving the overall design of the platform which includes a branded image for the platform as long as designed badges and improved interaction. We are also working with medical schools to spread the word of the utility of this platform.

V. CONCLUSIONS AND FUTURE WORK

We presented a web application that applies crowdsourcing methodologies together with a gamified approach to successfully translate a large *information structure* that, currently, is only available in one language. The platform is currently available in English and supports translating from English to any other language. Currently, spell checking and automatic translation methods are only available for Portuguese users. The next step is to improve the graphic design and make it available to a health professional community and by this be able to build a translated CHV. After these, the code will be open sourced. In the future we want to offer the features mentioned above to other languages and continue improving the quality of the results by, for example, making votes from users with more points have more value in the translation process. This application can be easily applied to different domains and other source languages. A working version of the platform is available at <http://irlab.fe.up.pt/p/healthtranslations>

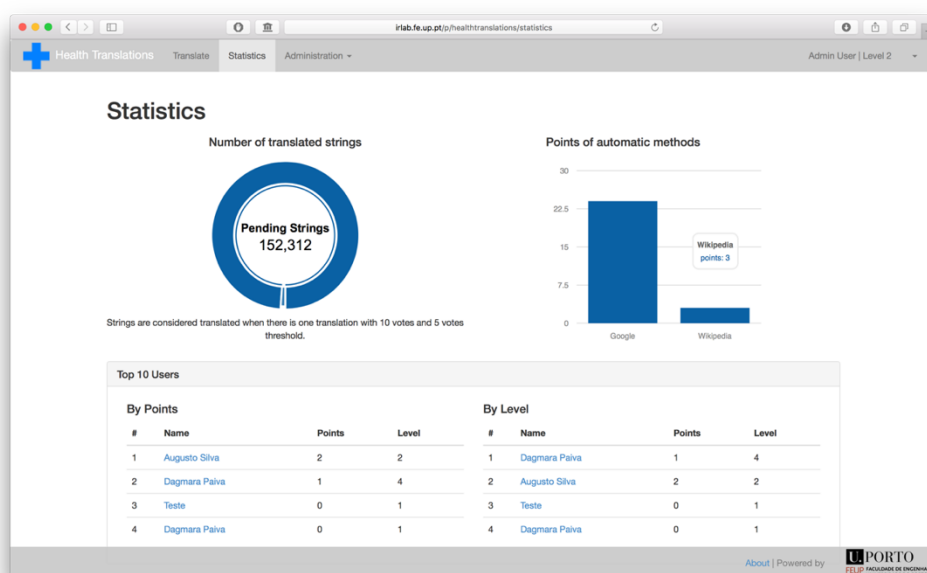


Figure 4. Statistics for administrators.

Search

Concept Id	CHV String	Best Translation	All Translations	Hide/Show
C0000892	acarus	acarus	View	Hide
C0000766	abnormal x rays	anormais x raios	View	Hide
	abnormal xray	anormais xray	View	Hide
	abnormal xrays	anormais raios x	View	Hide
	x-ray abnormal	x-ray anormais	View	Hide
C0000768	abnormal development	anormais desenvolvimento	View	Hide
	abnormalities congenital	anormalidades congénita	View	Hide
	abnormalities developmental	anormalidades desenvolvimento	View	Hide
	abnormality congenital	anormalidade congénita	View	Hide
	anomalies congenital	anormalias congénita	View	Hide
	anomalies fetal	anormalias fetal	View	Hide
	anomaly congenital	anormalia congénita	View	Hide
	birth defect	nascimento defeito	View	Hide
	birth defects	nascimento defeitos	View	Hide
	birthing defects	parto defeitos	View	Hide
	births defects	nascimentos defeitos	View	Hide
	congenital abnormalities	congénita anormalidades	View	Hide

About | Powered by

PORTO
FUP
FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Figure 5. Management of translations and strings.

References

- [1] University of Utah. Consumer Health Vocabulary Initiative. Visited on 01/10/2015. <http://consumerhealthvocab.org>.
- [2] Amazon Mechanical Turk. FAQs page. Visited on 12/10/2015. <https://www.mturk.com/mturk/help?helpPage=overview>
- [3] Transifex. Product page. Visited on 12/10/2015. <https://www.transifex.com/product/translate/>
- [4] Omar F. Zaidan and Chris Callison-Burch. 2011. Crowdsourcing translation: professional quality from non-professionals. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 1220-1229.
- [5] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining "gamification". In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11). ACM, New York, NY, USA, 9-15. DOI=10.1145/2181037.2181040 <http://doi.acm.org/10.1145/2181037.2181040>
- [6] Carla Teixeira Lopes, Cristina Ribeiro. Identification and Classification of Health Queries: Co-Occurrences vs. Domain-Specific Terminologies. International Journal of Healthcare Information Systems and Informatics (IJHISI), 9(3), 55-71, 2014. DOI=10.4018/ijhisi.2014070104.
- [7] Spence Green, Jeffrey Heer, and Christopher D. Manning. 2015. Natural Language Translation at the Intersection of AI and HCI. Queue 13, 6, pages 30 (June 2015), 13 pages. DOI=<http://dx.doi.org/10.1145/2791301.2798086>
- [8] Juho Hamari, Jonna Koivisto and Harri Sarsa. Does Gamification Work? A Literature Review of Empirical Studies on Gamification. 47th Hawaii International Conference on System Science. 2014.
- [9] Josceli Maria Tenrio, Ivan Torres Pisab. Consumer Health Vocabulary: A Proposal for a Brazilian Portuguese Language. MEDINFO 2015: eHealth-enabled Health. 2015.