

**WEBSITE LOCALIZATION OF KHAN
ACADEMY ARMENIA AND ITS IMPACT
ON ENGLISH INTO ARMENIAN
TRANSLATION INDUSTRY**

Capstone by Gayane Aramyan

American University of Armenia

2019



RESEARCH PAPER

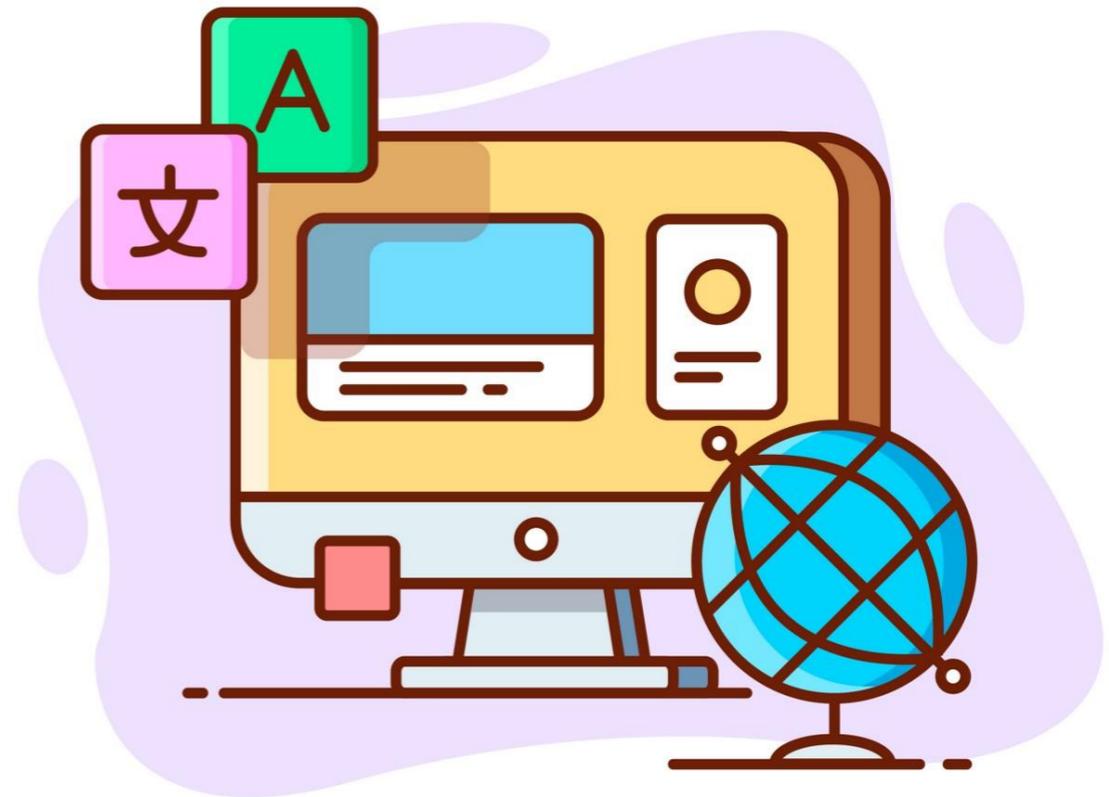
- This research is an examination of practices of localization of Khan Academy educational website from English language into Armenian. This study addresses what is the best way to translate, determine better processes of the software application development. The practical application goes through localization procedures of Khan Academy Armenia and evaluating its effectiveness.

METHODOLOGY

- ▶ The case study was a direct exploration of localization platforms related to Khan Academy project - crowdin.org platform and Khan Academy's localization tools developed by their engineers
- ▶ Investigation of accumulated translated data of Khan Academy localization from English language into Armenian (more than 3 million translated data)
- ▶ During the study two main areas examined: provided internationalization tools, technical solution and localization management system

KEY TERMS AND DEFINITIONS

- ▶ Internationalization (i18n)/globalization (g11n)
- ▶ Localization
- ▶ Translation memory (TM)
- ▶ Machine translation (MT)
- ▶ User experience (UX)
- ▶ Computer assisted translation (CAT)
- ▶ User interface (UI)



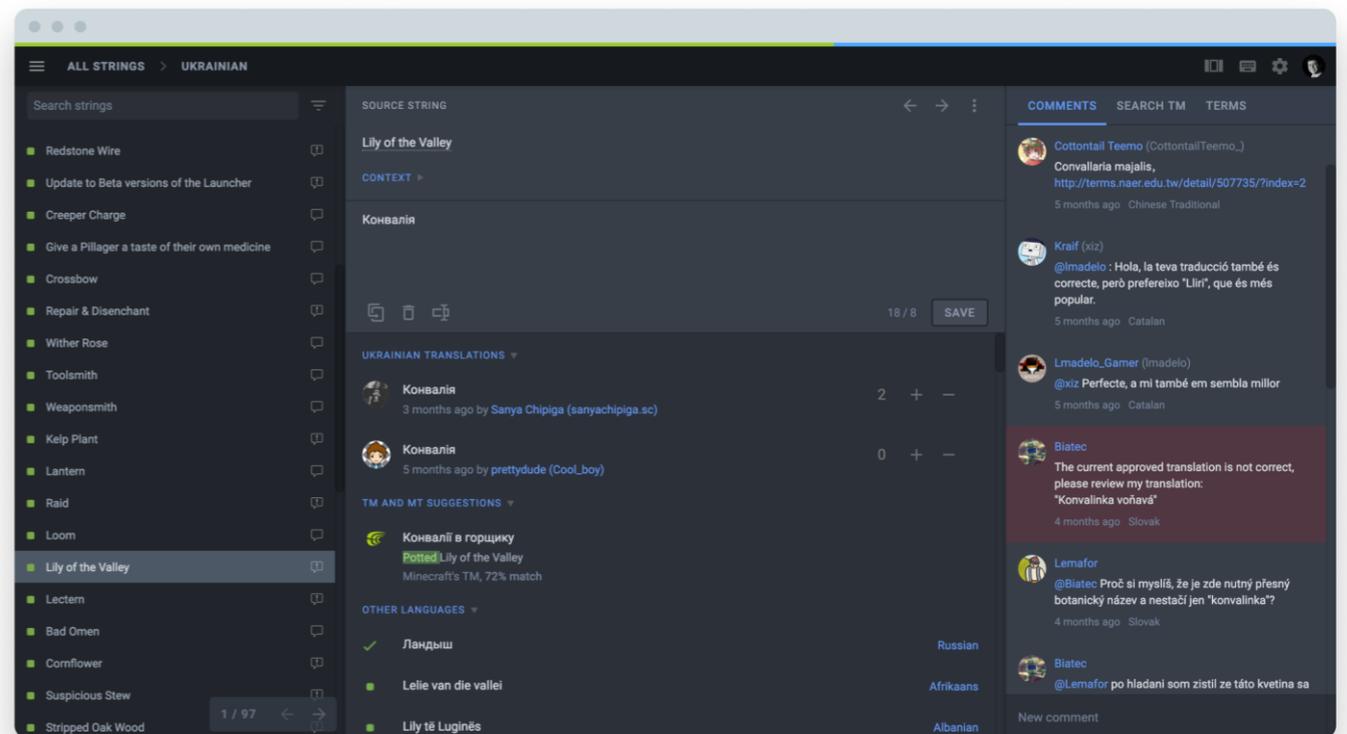
KHAN ACADEMY EDUCATIONAL PLATFORM



- ▶ Khan Academy platform is being localized in more than 36 languages throughout the world
- ▶ Khan Academy localization international team has more than 10,212 community of translators, 112 localization managers and annually it translates 57,650,000 words, out of which 36,126,000 is approved and published
- ▶ [crowdin.org](https://www.crowdin.org) is their outsourcing localization company

CROWDIN LOCALIZATION MANAGEMENT PLATFORM

➤ Crowdin is a localization management platform, which offers services to software, web site, apps, video games to localize their content into different languages.



EVALUATION OF LOCALIZATION OF KHAN ACADEMY ARMENIA

- 1. Localization tools** - provided services of [crowding.org](https://www.crowding.org). Understand the effectiveness of tools like automated translation, machine translation suggestion, translation memory and unified glossary, terminology for translators.
- 2. Localization management solutions** - providing high quality and cost-effective solutions for the project where translation and localization processes more efficiently can take place.

MACHINE TRANSLATION (MT) AND TRANSLATION MEMORY (TM) TOOLS OF CROWDIN

- ▶ In any source string translators can see automated translation suggestions and anything that was translated before hand. They also are able to see where the context will appear in the platform
- ▶ Translation memory can sometimes be a misleading tool, because it includes both approved and not approved data
- ▶ Machine translation can be helpful, but they always need to be double checked, because the system is not 100% accurate

MACHINE TRANSLATION (MT) AND TRANSLATION MEMORY (TM) TOOLS OF CROWDIN

e

The screenshot displays the Crowdin web interface for a project. At the top, the 'SOURCE STRING' section shows the text 'Done' in English, which is highlighted with a red box and an arrow pointing to the text 'Source string in English'. Below this, the 'CONTEXT' and 'EDIT' buttons are also highlighted with a red box and an arrow pointing to the text 'Where translation goes into'. The context text includes comments and a list of source tree entries for various languages. Below the source string, the 'Ավարտված է' (Completed) button is highlighted with a red box and an arrow pointing to the text 'Source string in Armenian'. The 'ARMENIAN TRANSLATIONS' section shows two completed translations, each with a profile picture, name, date, and approval status. The 'TM AND MT SUGGESTIONS' section is highlighted with a red box and an arrow pointing to the text 'Terminology and Machine translation suggestion'. It lists four suggestions, each with a green checkmark icon, the text 'Done!', and the source of the suggestion (Global TM or Official Khan Academy translations's TM) along with the match percentage.

SOURCE STRING

Done

Source string in English

CONTEXT EDIT

Where translation goes into

Ավարտված է

Source string in Armenian

ARMENIAN TRANSLATIONS

Ավարտված է
6 months ago by Margarit_Iskandaryan
approved 6 months ago by Margarit_Iskandaryan

Ավարտված է
2 years ago by Emma Kirakosyan (kirakosyanemma)

TM AND MT SUGGESTIONS

Done
Global TM, 100% match

Պատրաստ է
Global TM, 100% match

Կատարված է:
Done!
Official Khan Academy translations's TM, 80% match

Ավարտված է:
Done!
Official Khan Academy translations's TM, 80% match

Terminology and Machine translation suggestion

IMAGE LOCALIZATION EMBEDDED WITH TEXTS

- Crowdin does not support image localization processes if it is embedded with English texts.
- Even though Khan Academy's external platforms support image localization, however they must convert all image texts into source string. Remaking the image can cause proper image size, additional bug issues and in some instances irreversible translation errors.

Տարեթվերի վերաբերյալ առավել հաճախ տրվող հարցեր

✉ էլ. հասցե

Today is:

Gregorian: Wednesday, 26 December 2012
Mayan: Long count = 13.0.0.0.5; tzolkin = 9 Chicchan; haab = 8 Kankin
French: 6 Nivôse an 221 de la Révolution
Islamic: 12 Safar 1434
Hebrew: 13 Teveth 5773
Julian: 13 December 2012
ISO: Day 3 of week 52 of 2012
Persian: 6 Dey 1391
Ethiopic: 17 Takhsas 2005
Coptic: 17 Kiyahk 1729
Chinese: Cycle 78, year 29 (Ren-Chen), month 11 (Ren-Zi), day 14 (Xin-You)

Երբ է ստեղծվել մեր օրացույցը

Մենք գրում ենք 12/26/12թ. կամ 2012թ. դեկտեմբերի 26-ին, չորեքշաբթի: Ավանդաբար ընդունված է սա համարել որպես երկու հազար և տասներկու տարի (մի քանի տարվա տարբերություն) Հիսուս Քրիստոսի ծնունդից հետո: Բայց եթե Հիսուսն օրացույց օգտագործեր, ապա այն նման չէր լինի մեր օրացույցին:

IMAGE LOCALIZATION EMBEDDED WITH TEXTS

Տարեթվերի վերաբերյալ առավել հաճախ
տրվող հարցեր

✉ Էլ. հասցե

Today is:

Gregorian: Wednesday, 26 December 2012

Mayan: Long count = 13.0.0.0.5; tzolkin = 9 Chicchan; haab = 8 Kankin

French: 6 Nivôse an 221 de la Révolution

Islamic: 12 Safar 1434

Hebrew: 13 Teveth 5773

Julian: 13 December 2012

ISO: Day 3 of week 52 of 2012

Persian: 6 Dey 1391

Ethiopic: 17 Takhsas 2005

Coptic: 17 Kiyahk 1729

Chinese: Cycle 78, year 29 (Ren-Chen), month 11 (Ren-Zi), day 14 (Xin-You)

Երբ է ստեղծվել մեր օրացույցը

Մենք գրում ենք 12/26/12թ. կամ 2012թ. դեկտեմբերի 26-ին, չորեքշաբթի:
Ավանդաբար ընդունված է սա համարել որպես երկու հազար և տասներկու
տարի (մի քանի տարվա տարբերություններ) Հիսուս Քրիստոսի ծնունդից հետո:
Բայց եթե Հիսուսն օրացույց օգտագործեր, ապա այն նման չէր լինի մեր
օրացույցին:

REQUIREMENT OF CREATING ADDITIONAL LOCALIZATION PLATFORMS BY KHAN ACADEMY ENGINEERS

- Localization files of Crowdin does not reflect the content structure of Khan Academy, that is why Khan Academy's engineering team built additional platform for translators.
- It provides more data about translation and more efficient for task management

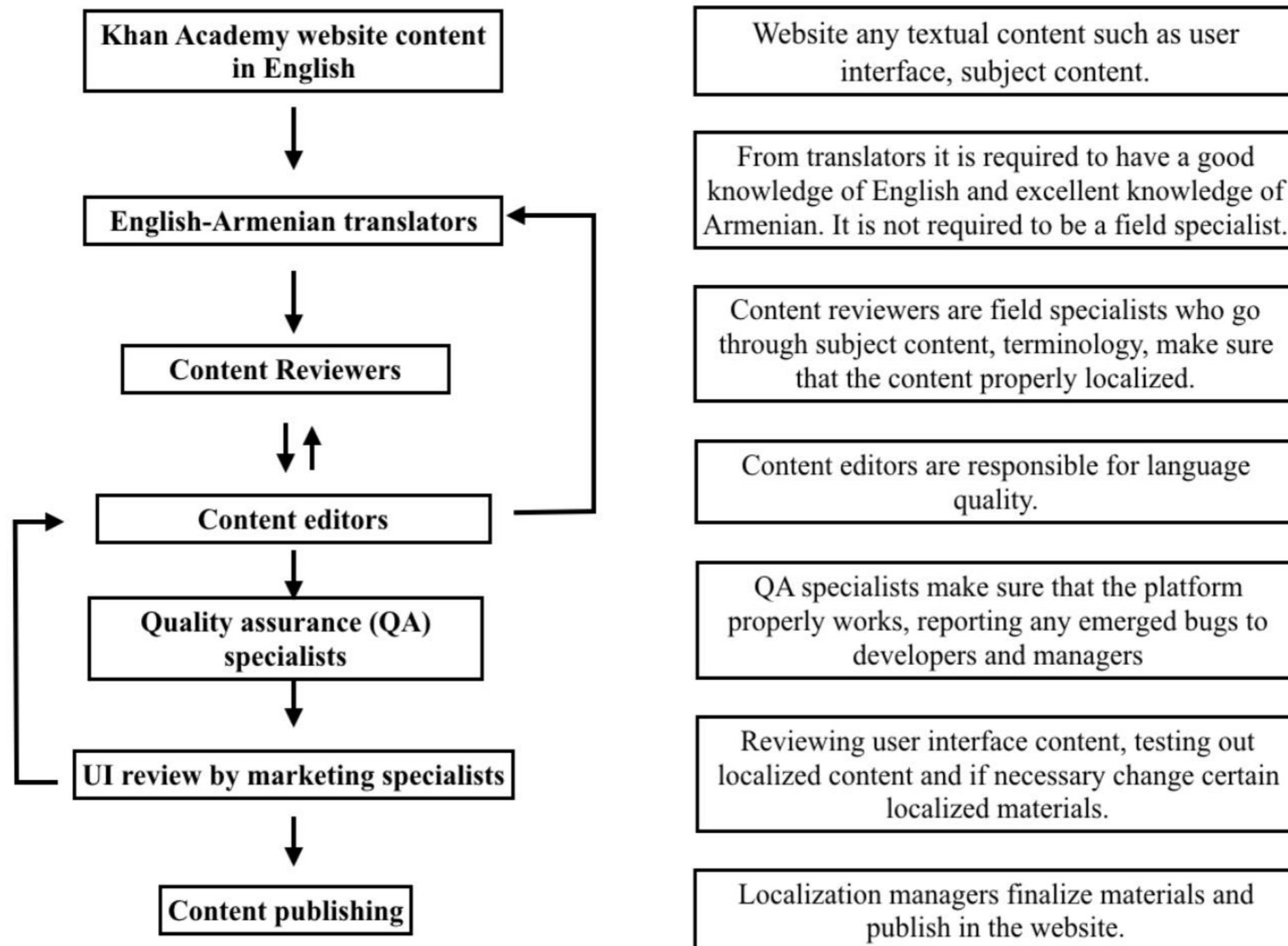


REQUIREMENT OF CREATING ADDITIONAL LOCALIZATION PLATFORMS BY KHAN ACADEMY ENGINEERS

- ▶ Khan Academy's translation dashboard is more structured and convenient to use, because in almost 18 million words data it was necessary to build localization platform which reflects the architecture of the software
- ▶ Two platforms provide machine translation, however both of them have high frequency errors, and in some cases it is better to avoid using them.
- ▶ Even though other additional platforms are built, still they must make sure that all the data synchronizes throughout all systems.

MODELING MORE EFFICIENT LOCALIZATION MANAGEMENT MECHANISMS

New model of Khan Academy Armenia's localization processes



KHAN ACADEMY ARMENIA - VIDEO LOCALIZATION

Video quality evaluation - Khan Academy Armenia Feedback content for subject video tutorials

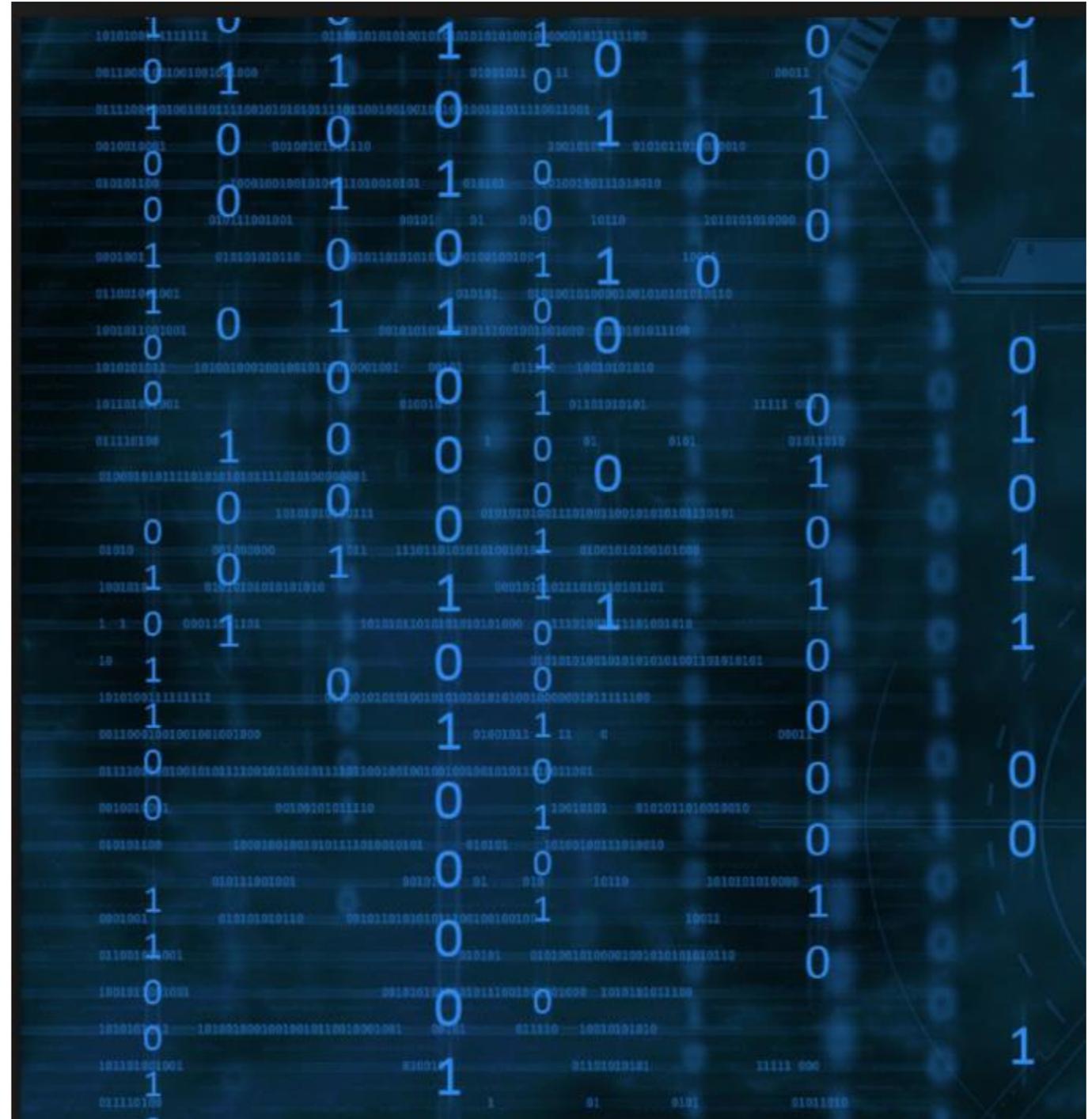
Content	Understandability of the concepts	The absence of conceptual mistakes	Color coding
Translation and adaptation	Accuracy of terms' translation	Communication of the message, the quality of adaptation	Consistency
Speech and diction	Natural and motivating diction	Dynamic diction (rhythmic, vigorous)	The absence of redundant words
Language	Literate writing	Literate speech	
Technical	Recording device problem	The quality of editing	Malfunctioning of tablet, improper color palette, arrangement problem, poor handwriting)

CONCLUSIONS FROM CASE STUDY

- For Khan Academy outsourcing localization is not cost-effective solution. As their platform data and structure has become more complex, they must make sure that localization always takes part in their product development processes.
- This way also engineers will find more efficient solutions related to technical abilities of the platform, in debugging processes, and bring more unified solutions.
- Besides technical support, localization management procedures must constantly measure their work efficiency, because with rapid technological developments translation team can miss out cost-effective ways of executing the project.

LIMITATIONS REGARDING THE RESEARCH

- Software is rapidly developing industry, thus any innovation with it brings new challenges to localization.
- Machine translation industry has promising future, however its benefits for many languages may not be useful in the near future.



REFERENCES

- Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge.
- Chan, S. (2015). *The Routledge encyclopedia of translation technology*. London: Routledge.
- Cronin, M. (2013). *Translation in the digital age*. London: Routledge.
- Deitsch, A., & Czarnecki, D. (2001). *Java internationalization*. Sebastapol: O'Reilly.
- De la Cova, E. (2016, November). Translation Challenges in the Localization of Web Applications. *Sidebar* 27, 135-166.
- Dunne, K. J., & Dunne, E. S. (2011). *Translation and localization project management: The art of the possible*. Amsterdam: John Benjamins Pub.
- Esselink, B. (2000). *A practical guide to localization (pp. 25-54)*. Amsterdam: John Benjamins.

REFERENCES

Franklin, D. (2018). *Megatech: Technology in 2050*. London: The Economist.

Frederking, R. E., & Taylor, K. B. (2004). *Machine translation: From real users to research: 6th conference of the Association for Machine Translation in the Americas, AMTA 2002, Washington, DC, USA, September 28 - October 2, 2004: Proceedings*. Berlin: Springer.

Globalization and Localization Association (GALA) (2012). *About the Localization Industry*. Retrieved from <http://www.gala-global.org/about-localization-industry>.

Jiménez-Crespo, M. A. (2011, January). To adapt or not to adapt in web localization: a contrastive genre-based study of original and localized legal sections in corporate websites. *The Journal of Specialised Translation*, 15, 2-27.

Jiménez-Crespo, M. A. (2013). *Translation and web localization*. London: Routledge.

LeBlanc, M. (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *Translation and Interpreting* 5 (2), 1–13.

Malmkjær, K., & Windle, K. (2011). *The Oxford handbook of translation studies*. Oxford: Oxford University Press.

REFERENCES

Pym, A. (2014). *Exploring translation theories*. London: Routledge.

Schwieter, J. W., & Ferreira, A. (2017). *The handbook of translation and cognition*.

Chichester: Wiley-Blackwell.

Suojanen, T., Tuominen, T., & Koskinen, K. (2015). *User-centered translation (pp.13-29)*.

London: Routledge.

Venuti, L. (2013). *Translation changes everything: Theory and practice*. London: Routledge.

Wilks, Y. A. (2010). *Machine Translation: Its Scope and Limits*. New York: Springer.

THANK YOU