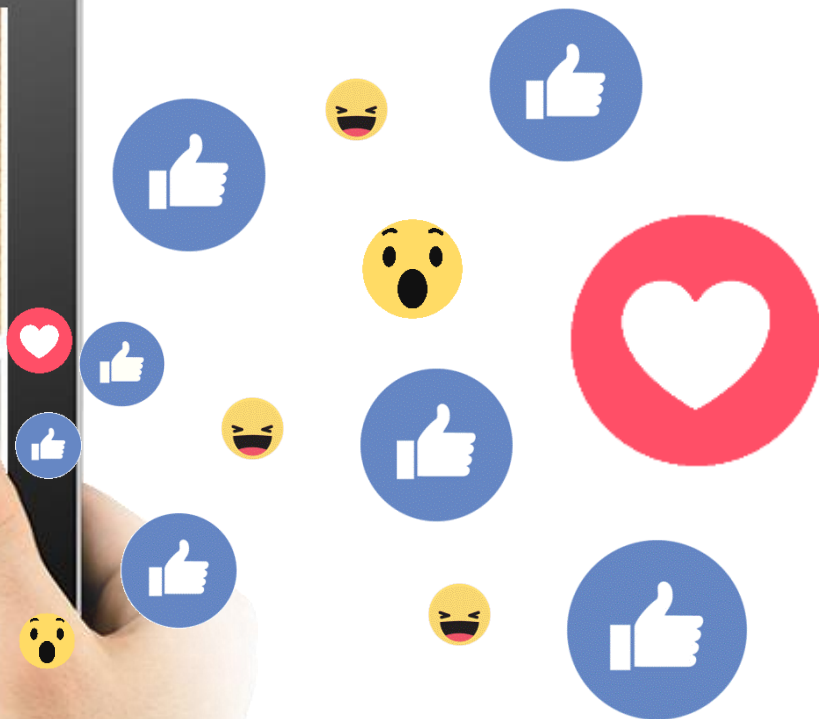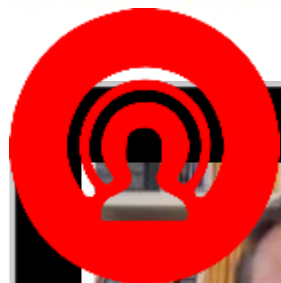# Webinar-2019

Localization Engineering: The Magic Secret to Making Translation Faster, Cheaper and Better Quality

Oksana Tkach

April 25, 2019

The next summer is for UTICamp
July 22-28

Early Bird Registration

https://utic.eu/en/o-konferentsii

About the conference

19.03.2019   9 good reasons to register for UTICamp-2019 right now!

Register to UTICamp

1. Take part in the biggest industry event in Ukraine

Since 2013 our conference has been gathering together hundreds of participants from many countries around the world, and dozens of speakers have shared their experience, knowledge and unique ideas. You too can become a part of this event and tap into the source of translation knowledge known as UTICamp! Learn more about the 2013, 2014 and 2016 conferences in the Events Archive section.

2. Select a convenient time to participate

We are holding the event over a whole week. You can opt for full-week participation or choose one of the Masters Village or Three Tracks segments. The "first shift" of the conference on July 23 and 24 will be dedicated to masters - professionals, experts, creators, and gurus, while the Three Tracks days will feature traditional presentations in three parallel streams — The Art of Translation, The Translation Business and Translation Technology and Localization. Find out more about what to expect at the conference on July 22—28.

The preliminary conference program can be found here.

3. Combine conference participation with a family vacation and work

You can attend our conference with your family. While you listen to presentations and spend time with colleagues from the industry, your children can play with the camp entertainers in a separate play zone, where a special fun program will be organized for them. And a guest ticket can be bought for adults who do not plan to attend the workshops and presentations.

Do you want to participate in the conference without stopping your work? Wi-Fi and a coworking area will be at your disposal.

4. Listen to presentations by translation industry experts

UTICamp is an international conference where you can meet outstanding speakers and translation industry experts: people with unique experience and vivid ideas.

info@utic.eu

UKRAINIAN
TRANSLATION
INDUSTRY
CAMP
UTIC

July 22–28, 2019

RU   UK

About the conference   Conference news   Sponsors   Video   Contacts   Past events

Sponsors

We are glad to see our standing partners among UTICamp-2019 sponsors. Thank you for supporting our conference once again!

With the prominent coverage of UTICamp events, you will strengthen your position in the market, expand your network of partners and vendors, further your brand and become an influencer for the industry.

Learn more about the sponsorship opportunities and make your contribution!

Iryna Vizir
UTICamp CMO

➢ **Oksana Tkach** studied linguistics and worked as a translator and localization specialist at Linguistic Centre and BlaBlaCar.

➢ During this time, she shared her experiences in the localization industry as a speaker at the European Commission's Translating Europe Forum and UTICamp 2016.

➢ She later went on to study Data Science and Computational Linguistics, and worked as a data scientist before moving on to start her own company called Metamova.

➢ Currently, Metamova is mainly working with Lionbridge to provide localization engineering and development services, but is also collaborating with other LSPs and companies looking to automate their processes and leverage their language data using Natural Language Processing, or to create NLP products.

➢ Oksana and her team are passionate about the localization

What we'll talk about

1. Why localization engineering is important

2. Localization engineering in production

3. Bigger picture: automate, automate, automate!

4. But this is not all! Computational linguistics
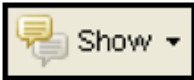
5. Taking it all further: Data Science

1. Why localization engineering is important

Why it's important
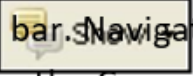
HOW TO...		
Action	**Adobe Reader version 9**	**Adobe Reader version X and XI**
Review	To review your changes, click on the 'Show' button [Show ▾] on the Commenting tool bar. Choose 'Show Comments List'. Navigate by clicking on a correction in the list. Alternatively, double click on any mark-up to open the commenting box.	Your changes will appear automatically in a list below the Comment tool bar. Navigate by clicking on a correction in the list. Alternatively, double click on any mark-up to open the commenting box.
Undo/delete change	To undo any changes made, use the right click button on your mouse (for PCs, Ctrl-Click for the Mac). Alternatively click on 'Edit' in the main Adobe menu and then 'Undo'. You can also delete edits using the right click (Ctrl-click on the Mac) and selecting 'Delete'.	To undo any changes made, use the right click button on your mouse (for PCs, Ctrl-Click for the Mac). Alternatively click on 'Edit' in the main Adobe menu and then 'Undo'. You can also delete edits using the right click (Ctrl-click on the Mac) and selecting 'Delete'.

HOW TO...		
Action Adobe Reader version 9 Adobe Reader version X and XI		
Review	To review your changes, click on the 'Show' below the Comment tool bar. Navigate by clicking on a correction in the list. Alternatively, bar. Choose 'Show Comments List'. Navigate by clicking on a correction in the list. Alternatively, double click on any mark-up to open the commenting box.	Your changes will appear automatically in a list
Undo/delete change	To undo any changes made, use the right click button on your mouse (for PCs, Ctrl-Click for the Mac). Alternatively click on 'Edit' in the main Adobe menu and then 'Undo'. You can also delete edits using the right click (Ctrl-click on the Mac) and selecting 'Delete'.	To undo any changes made, use the right click button on your mouse (for PCs, Ctrl-Click for the Mac). Alternatively click on 'Edit' in the main Adobe menu and then 'Undo'. You can also delete edits using the right click (Ctrl-click on the Mac) and selecting 'Delete'.

https://nuts-about-needlepoint.com/scattered-needlepoint-areas-back/

https://nuts-about-needlepoint.com/scattered-needlepoint-areas-back/

This is why you need some engineering magic!

Translation process

PM accept file → DTP (PDF) → Engineering → Translation process → Engineering → DTP (PDF) → PM deliver file
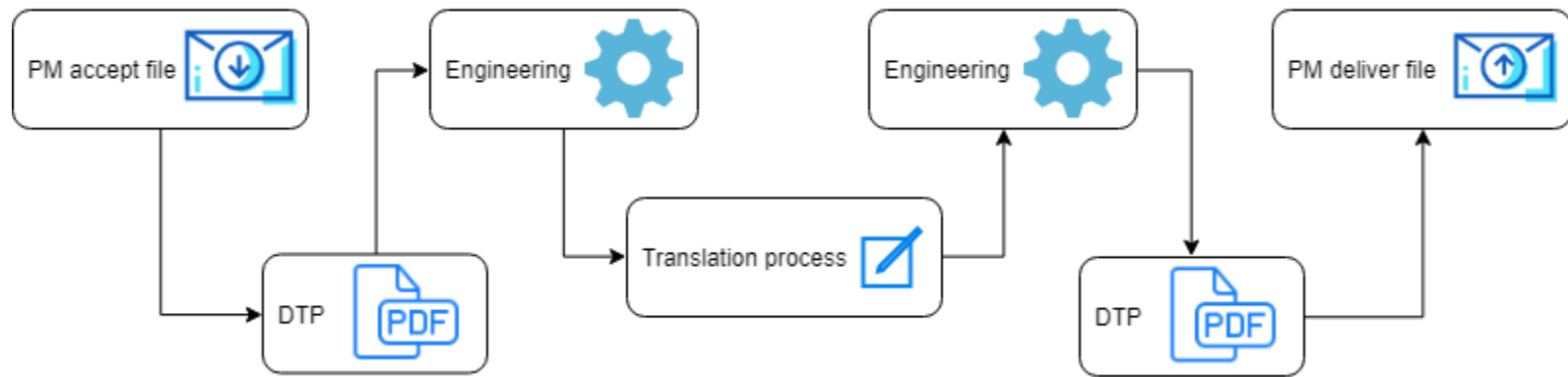
2. Localization engineering in production

Production task: provide the translator with only the text they have to translate; nothing less, nothing more.

Production task: provide the translator with only the text they have to translate; nothing less, nothing more.

Not as easy as it sounds.

Production

UKRAINIAN
TRANSLATION
INDUSTRY
CAMP

UTIC

July 22–28, 2019

RU   UK

Conference news   About the conference   Prices   Program   Sponsors   Video   Contacts   Past events   Attendees

```html
<li class="home"><a href="https://utic.eu/en/novosti" target="_self">Conference news</a></li>

<li><a href="https://utic.eu/en/o-konferentsii" target="_self">About the conference</a><ul><li><a href="https://utic.eu/en/o-konferentsii/9-preimushchestv-zaregistrirovatsia-na-uticamp-2019-uzhe-seichas" target="_self">Participate in UTICamp-2019</a></li>
<li><a href="https://utic.eu/en/mesto-provedeniia" target="_self">Conference venue</a></li>
<li><a href="https://utic.eu/en/tseny-i-usloviia-uchastiia-2019" target="_self">Prices and terms of participation</a></li>
<li><a href="https://utic.eu/en/chto-vas-ozhidaet-na-konferentsii-2228-iiulia" target="_self">What to expect at the conference</a></l
<li><a href="https://utic.eu/en/na-uticamp-s-semei" target="_self">Attending UTICamp with your family</a></li>
<li><a href="https://utic.eu/en/gde-ostanovitsia" target="_self">Where to stay</a></li>
<li><a href="https://utic.eu/en/sightseeing-at-uticamp-2019" target="_self">Sightseeing at UTICamp-2019</a></li>
</ul></li>

<li><a href="https://utic.eu/en/tseny-i-usloviia-uchastiia-2019" target="_blank">Prices</a></li>

<li><a href="https://utic.eu/en/program" target="_self">Program</a></li>

<li><a href="https://utic.eu/en/sponsors" target="_self">Sponsors</a></li>

<li><a href="https://utic.eu/en/video/videodoklady" target="_self">Video</a><ul><li><a href="https://utic.eu/en/video/videodoklady" t
"_self">Presentations</a></li>
<li><a href="http://webinar.utic.eu/">Webinars</a></li>
<li><a href="https://utic.eu/en/video/treningi" target="_self">Trainings</a></li>
</ul></li>

<li><a href="https://utic.eu/en/contact_us" target="_self">Contacts</a></li>
```

Production

Source	Target
Conference news	
About the conference	
Participate in UTICamp-2019	
Conference venue	
Prices and terms of participation	
What to expect at the conference	
Attending UTICamp with your family	
Where to stay	
Sightseeing at UTICamp-2019	
…	

Production

- XLIFF

- PO

- CSV

- JSON

- PROPERTIES

…

Production

```xml
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<xliff xmlns="urn:oasis:names:tc:xliff:document:1.2" version="1.2">
    <file>
        <body>
            <trans-unit>
                <source>
                    Translate me, please!
                </source>
                <target>
                    Переклади мене, будь ласка!
                </target>
            </trans-unit>
        </body>
    </file>
</xliff>
```

```xml
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<xliff xmlns="urn:oasis:names:tc:xliff:document:1.2" version="1.2">
    <file>
        <body>
            <trans-unit>
                <source>
                    <![CDATA[
                        <h6 style="color:red;" align="right">Translate me, please!</h6>
                    ]]>
                </source>
                <target>
                    <![CDATA[
                        <h6 style="color:red;" align="right">Переклади мене, будь ласка!</h6>
                    ]]>
                </target>
            </trans-unit>
        </body>
    </file>
</xliff>
```

Get our 50% off promo code and save over $80 on your second order!

Get our 50% off promo code and save over $80 on your second order!

```
"Get our %s%% off promo code and save
over $%s on your second order!" %
("50", "80")
```

Get our 50% off promo code and save over $80 on your second order!

```
"Get our %s%% off promo code and save
over $%s on your second order!" %
("50", "80")
```

Source	Target
Get our <%s>% off promo code and save over $<%s> on your second order!	

APRIL 11, 2019
#140 The Roman Mars Mazda Virus
by REPLY ALL
LISTEN NOW

Roman Mars' 99% Invisible podcast

https://gimletmedia.com/shows/reply-all/brh8jm/140-the-roman-mars-mazda-virus

Encoding Madness

SourceLang	SourceText	TargetLang	TargetText
en-US	What can be reported through this channel?	ja-jp	このポータルを通じてどのようなことを報告できますか。

Encoding Madness

SourceLang	SourceText	TargetLang	TargetText
en-US	What can be reported through this channel?	ja-jp	このポータルを通じてどのようなことを報告できますか。

SourceLang	SourceText	TargetLang	TargetText
en-US	What can be reported through this channel?	ja-jp	???????????????????????????

Encoding Madness

SourceLang	SourceText	TargetLang	TargetText
en-US	What can be reported through this channel?	ja-jp	このポータルを通じてどのようなことを報告できますか。

SourceLang	SourceText	TargetLang	TargetText
en-US	What can be reported through this channel?	ja-jp	???????????????????????????

SourceLang	SourceText	TargetLang	TargetText
en-US	What can be reported through this channel?	ja-jp	ã"ãªãƒãƒ¼ã‚¿ãƒ«ã'é€šã˜ã¦ã©ãªã‚ˆã†ãªã"ã'±å'Šã§ãã¾™ã‹ã€

YO DAWG, I HEARD YOU LIKE UTF-8 CHARACTER ENCODING

SO I WENT AHEAD AND ENCODED YOUR LATIN_1 IN MY UTF8 AS ASCII/MYISAM CHARACTERSET

memegenerator.net

3. Bigger picture: automate, automate, automate!

Workflows and process architecture task:

We want you to upload a file and receive the exact same file in a different language with minimum possible human intervention.

Process Automation

Task Sequence: Custom ▾   Task Sequences...

Details

Available Tasks:

Convert to Translatable Format
Word Count
Copy to Target Languages
WIP Report
Verify Files
Update Project Translation Memories
Update Main Translation Memories
Translation Count
Pre-translate Files
Update from Bilingual Review
Create Regulator Review Bundle
Update from Reviewed Target File (Retrofit)
Pseudo-translate
Populate Project Translation Memories
Apply PerfectMatch
Translation Quality Assessment
Export for Bilingual Review
Analyze Files
Export files
Generate Target Translations

Add >

< Remove

Selected Tasks:

Convert to Translatable Format
Copy to Target Languages
Analyze Files
Pre-translate Files
Update Project Translation Memories
Export files

AUTOMATE

ALL THE THINGS

Computers can't distinguish between language and code, language and placeholder, translatable and untranslatable text.

Process Automation

```
{

   "235234": {

      "header][0][value": "<div class=\"class_name\"><p>Translate me,
please!  <\/p><\/div>"

   },

   "235235": {

      "body][0][value": "<h2>Translate me, please!<\/h2><p>Translate
me, please!<\/p>"

   }
}
```
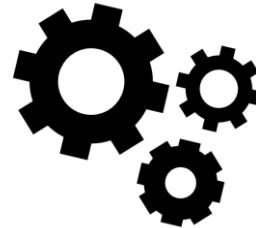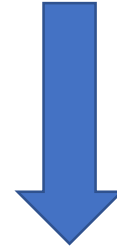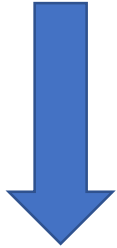
```
{
    "235234": {
        "header][0][value": "<div class=\"class_name\"><p>Translate me, please!  <\/p><\/div>"
    },
    "235235": {
        "body][0][value": "<h2>Translate me, please!<\/h2><p>Translate me, please!<\/p>"
    }
}
```

Process Automation

```
{
    "235234": {
        "header][0][value": "<div class=\"class_name\"><p>Translate me, please!  <\/p><\/div>"
    },
    "235235": {
        "body][0][value": "<h2>Translate me, please!<\/h2><p>Translate me, please!<\/p>"
    }
}
```

```
{
    "235234": {
        "header][0][value": "<div class=\"class_name\"><p>Translate me,
please!  <\/p><\/div>"
    },
    "235235": {
        "body][0][value": "<h2>Translate me, please!<\/h2><p>Translate
me, please!<\/p>"
    }
}
```

Less source file types – more automation

3. But this is not all! Computational linguistics

Word count

Word count

How do we decide what is a word?

Word count

How do we decide what is a word?

Is "forty-seven" one or two words?

Word count

How do we decide what is a word?

Is "forty-seven" one or two words?

Is "he's" or "I'm" one or two words?

Word count

How do we decide what is a word?

Is "forty-seven" one or two words?

Is "he's" or "I'm" one or two words?

Is "Geschwindigkeitsbegrenzung" one word for a translator?

Word count

How do we decide what is a word?

Is "forty-seven" one or two words?

Is "he's" or "I'm" one or two words?

Is "Geschwindigkeitsbegrenzung" one word for a translator?

北京在中国北方；广州在中国南方。 Splitting Chinese?

Sentence splitting

Sentence splitting

How do we decide what is a translation unit?

Sentence splitting

How do we decide what is a translation unit?

This was first said by Shakespeare: "To thine own self be true."

Sentence splitting

How do we decide what is a translation unit?

This was first said by Shakespeare: "To thine own self be true."

There is mounting evidence of global warming; of course, some people will never believe it.

Sentence splitting

How do we decide what is a translation unit?

This was first said by Shakespeare: "To thine own self be true."

There is mounting evidence of global warming; of course, some people will never believe it.

<p>This is an example of a two sentence paragraph. It contains two sentences, but for some reason it's one translation unit.</p>

Sentence splitting

TM:

TM match:

This is an example of a two sentence paragraph. It contains two sentences, but for some reason it's one translation unit.

50% fuzzy match

This is an example of a two sentence paragraph.

50% fuzzy match

It contains two sentences, but for some reason it's one translation unit.

Non-translatable or substitutable text

Non-translatable or substitutable text

Dates: 11/02/2019

Links: https://utic.eu/uk

Emails: oksana@metamova.com

Addresses: 445 Mount Eden Road, Mount Eden, Auckland

Names: John Doe, CTO and co-founder

Non-translatable or substitutable text

Solutions:

      Regular expressions

      Gazetteers

Glossary check

Find glossary terms automatically – TF-IDF:

Find words that are used rarely in general, but frequently in this document/project/job.

Glossary check

Find glossary terms automatically – TF-IDF:

Search Word: find

wherever you go to on the english coast you **find** a number of batl

alice_in_wonderland.txt

find was found 21 times.
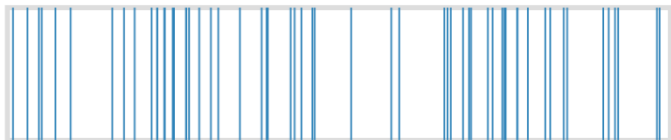
heaven i am going to morrow where i shall **find** a man who has no

pride_and_prejudice.txt

find was found 58 times.

http://vallandingham.me/concordance_plot/

Glossary check

Reduce the amount of false positives using lemmatizers:

English	Ukrainian
term	термін

Let's make term check more advanced with lemmatizers!	Давайте покращимо перевірку термінів з лематизаторами!

Glossary check

Reduce the amount of false positives using lemmatizers:

English	Ukrainian
term	термін, терміна, терміну, термінові, терміном, терміні, терміни, термінів, термінам, термінами, термінах

Let's make term check more advanced with lemmatizers!	Давайте покращимо перевірку термінів з лематизаторами!

5. Taking it all further: Data Science

- Neural Machine Translation
- Big data
- Data silos and sharing
- Using client TMs to fine-tune NMT
- Why machine translation will always need human translators

Where to find our videos

All video of speeches and webinars on our website:
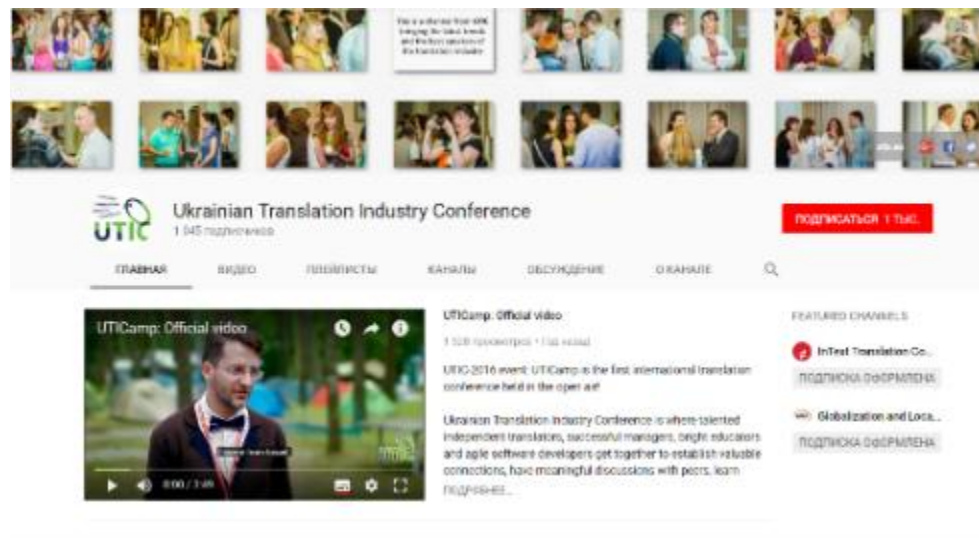
www.youtube.com/user/UTIConf

Join us on social media