

Trends in Translation Technologies

TM Conference
Krakau 2010



Angelika Zerfass
Consulting and Training
zerfass@zaac.de

Translation Memory

- Going from desktop application to server application to web application
- Inclusion of project management (interfaces with existing project management tools) or adding own PM components
 - Although internal PM features are rather project creation components than real project management systems
- Inclusion of workflow capabilities (file handling in preparation for translation and interaction with translators / clients) with automation
- Wider adoption of standard file formats in addition to TMX (XLIFF, TBX)

Translation Memory

- Translation Memories / review process going online (some examples)
 - Ontram (commercial TM tool for translation and review process online)
 - Wordfast Anywhere (free tool to build your own TM online as well as connect to the VLTM – very large translation memory from donated TMs, which is public)
 - Google Translator Toolkit (translation and review happen online, you upload documents and TMs, if nothing is found in a TM, Google Translate for automatic translation is used)
- Translation Memory connecting to CMS; CMS tools running part of the file handling, giving out translation tool friendly files and settings files
- VLTM (very large public TMs) online (for Wordfast users)

Editing Environment

- Traditional tools tend to move away from Word as an editor and create their own editing environment (desktop or browser-based)
- General move towards browser-based editing
 - Ontram <http://www.andrae-ag.de/index.php> (translation management and review tool online)
 - Amanuens <http://amanuens.com/> (editor for SW localization)
- New tools tend to integrate into Word first
 - WebWordSystems (online TM)
 - Jive Fusion (integration into Office products)
 - MultiTrans: online access via Word to the bi-text corpus

Wordfast Anywhere

Wordfast Anywhere - wordfast_translation memory software_tm & cat tools.doc - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http:// /anywhere/ Live Search

Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten

Wordfast Anywhere - wordfast_ ... x

File TM Term Tools ?

WORDFAST anywhere [logout](#)

Translation memory EN-US>DE-DE : 1868074 public TUs, 0 private TUs

paste
upload

Analysis report

tm & cat tools.doc	segments	words	char.	%
Analogy				
Repetitions	11	73	447	9%
100%	0	0	0	0%
95% - 99%	0	0	0	0
85% - 94%	0	0	0	0
75% - 84%	0	0	0	0
0% - 74%	101	731	4275	90
Total	112	804	4722	

(character count includes spaces)
(148 internal tags found. Tags are included in this report, accounted for as 1 tag=1 word; 1 tag=0 character.)

Close

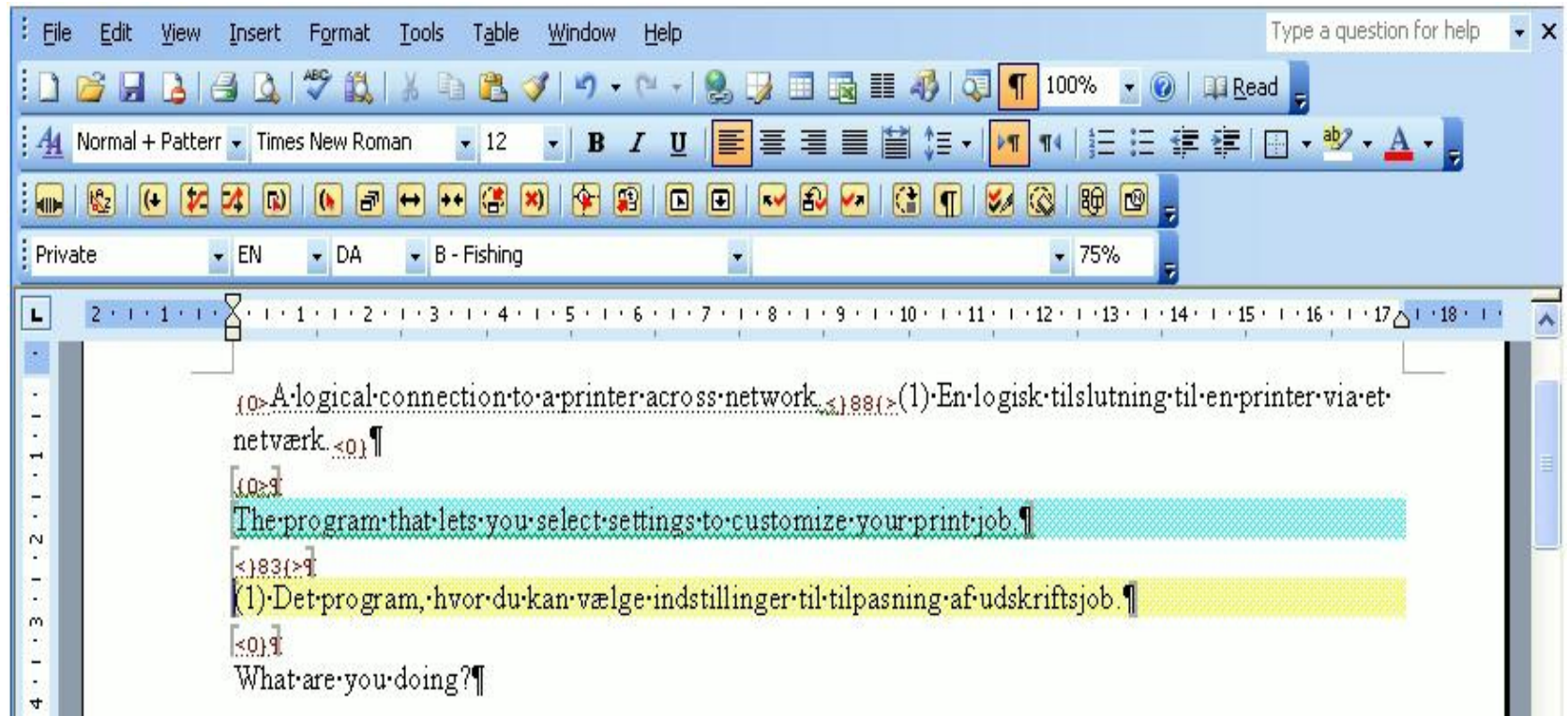
Learn More

<1>Buy Wordfast <2>Start saving time, money, and effort from the next word you translate<3>

Wordfast Product Briefs

<1>Wordfast Pro<2>Wordfast Classic<3>Wordfast Server <4>The professionals overview of features.

0 2/112 EN-US>DE-DE Bilingual



WebWordSystem – Online TM connecting to Word

Size of the Segments

- Traditional, database-driven translation tools add context information (surrounding segments, structural information) to produce better than 100% matches
- Translation memory tool with parallel texts instead of isolated segment pairs as repository add TM functionality to deal with single segment matches during translation (MultiTrans: adding a sentence based TM to the bi-text corpus)

Size of the Matches

- Matching for smaller elements than whole segments
 - Déjà Vu: assembling of text fragments using example based machine translation approach
 - memoQ: Fragment translation of known segment parts
 - Trados Studio: Autosuggest, list of phrase pairs created out of a statistical analysis of existing TMs
 - MultiTrans: WordAlign, using term extraction technology to find out how a term/phrase has been translated and what different translations there are in what frequency
- Calculation of internal matches (matches from your own translations not from the initial TM) or analysis of the translated document and statistics on the match rates that were achieved during translation

Calculation of Matches

- Calculation of internal matches
 - matches from your own translations not from the initial TM
- Analysis of the translated document (after translation)
 - the translation tool stores the match rates you got during translation

3 Trends in TM tools

• 3 Trends for TM tools

- Existing, sophisticated translation memory tools will not add own functionalities (or only little) but will connect more closely with other tools (project management, content management...)
- Content creation tools adding their own small translation components (MadCap Lingo)
- Online translation tools use public online TMs and private online TMs (WebWordSystems, Wordfast Anywhere)

• Open Source Tools

- http://www.folt.org/index.php?option=com_content&task=view&id=35&Itemid=51
- FOLT Open Source Translation Memory
- Open TM2 (previously Translation Manager, IBM)

Translation Tools for Free?

● Renato Beninatto, Milengo

- **Role and Cost of Technology:** Translation technology tools like translation management systems and desktop translation memory software are unusually expensive for the value that they provide. As Chris Anderson points out in his book "Free," software and content want to be free. What creates value are the services provided around technology, not the technology itself. So, I have predicted that Translation Memory technology will be free or become irrelevant by 2015.
- <http://www.glggroup.com/News/Some-Trends-in-the-Translation-and-Localization-Industry-49440.html>

● What do you think?

Translation Memory and Machine Translation

- Round and round they go...
 - TM repositories were actually developed out of the storage unit of a machine translation system
 - TM and MT go into different direction
 - MT starts to use not only rule-based and statistics-based systems but also example-based translation
 - Where to get many good examples of translations so that the MT system can learn how to translate certain phrases?
 - Use the content of Translation Memory systems
 - MT systems are fed by huge amounts of segment pairs from translation memories
 - MT and TM complement each other

Machine Translation

- It is here to stay. It will find its way into the translation process together with controlled language (for creation of the source language material) and post-editing (a new skill set)
 - Every editing tool will be able to connect to online MT (like the translate function in MS Word)
 - It will help translate material that otherwise would not get translated at all (for example "online support pages" – we are not talking about the quality here, just the availability)

Translation Memory and MT

some examples

- Déjà Vu uses an example based machine translation component to "repair" fuzzy matches from the TM
- Across, SDL – Language Weaver
- SDL Studio – Google Translate
- OmegaT – Google Translate
- Multitrans – Systran
- GlobalSight – Prompt
- API allows to connect TM systems with online translation services

Terminology

- As terminology management tools themselves are rather complex, online platforms are created to share terminology and even collaborate on it (between vendor and client)
 - many vendors create their own, as they do not want to use the platform of a competitor
- Free terminology offerings
 - Microsoft Terminology, Termium (EN-FR) linguistic database of Canada's Government, Linguee, free German – English dictionary with context sentences, from alignment of webpages
- Collaborative tools (wikis, blogs, project specific groups/forums/chats created by the server translation tool...) and voting systems in crowdsourcing allow a more interactive work on terminology
- TBX as a (ISO) standard for storing and exchanging term base data catches on (especially because of TBX Basic)

Online Terminology Management

Here are some examples, how such platforms can look...

- <http://demo.termexplorer.com/> (Termexplorer)
- <http://www.dog-gmbh.com/index.php?id=100> (Lookup)
- <http://termwebdemo.termweb.se/termweb/app>
(TermWeb)
- <http://www.evoterm.net/en/Demo.aspx> (evoterm)
- <http://www.i-term.dk/> (i-term)

- Abblaseeinrichtung
- abblasen
- Abdeckblech
- Abdeckhaube
- Abdeckkappe
- Abdeckscheibe
- Abdeckung
- Abdruckgewinde
- Abfluss
- Abfrage
- Abkühlung
- Ablagefach
- Ablageplatte
- Ablageplatz
- Ablassschraube
- Ablassventil
- Ablaufanzeiger
- Ablaufprogramm
- Ablaufrinne
- Ablaufspeicher


 Ihre Suche ergab 2733
 Einträge von 5466 im
 Wörterbuch



Eintragsnummer 63 In Werkliste

Quelle: Wikipedia

Definition Der Abziehstein wird für den Feinschliff und das Abziehen von Werkzeugschneiden aus Stahl verwendet.



Deutsch

Abziehstein

Verwendung Erlaubt

Beispiel 40: Eventuelle Dellen oder Gräte müssen mit einer Feile oder einem Abziehstein geglättet werden um eine Beschädigung der hochwertig gefertigten Tischoberfläche zu vermeiden.

Status Freigegeben

Vorgeben am 25.02.2009

Geändert am 17.03.2009

Geändert durch AZ

Englisch

oilstone

Verwendung Erlaubt

Beispiel 25: Possible indentations or burrs must be smoothed out with a file or an oilstone in order to avoid damaging the high-quality table surface.

Status nicht freigegeben

Vorgeben am 25.02.2009

Geändert am 17.03.2009

Geändert durch AZ

grindstone

Verwendung Verboten

Status Freigegeben



Source: English

Go

- Search
- Add concept
- References
- Import/Export
- Diagrams
- Layout
- Lists
- Users
- Tools
- Statistics
- Validate
- Restrictions

- About i-Term
- Contact
- Help
- Log off

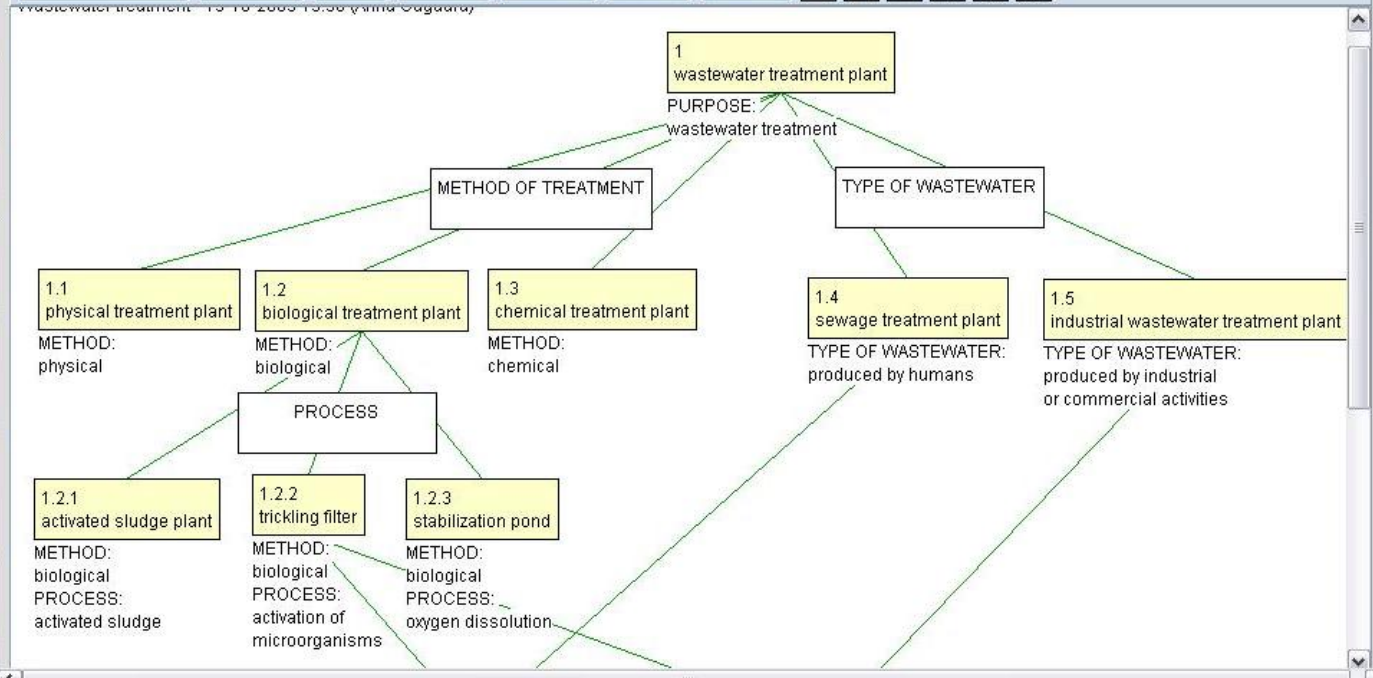
Database: Demo 3.4.9
 User: Anna Odgaard
 Authorization: Administrator
 Logged on: 13-10-2009 15:11
 Change password

- [biological treatment](#)
- [chemical treatment](#)
- [industrial wastewater treatment](#)
- [industrial wastewater treatment plant](#)
- [physical treatment](#)
- [sewage treatment plant](#)
- [sewage treatment works](#)
- [sewage trickling filter](#)
- [stabilization pond](#)
- [trickling biofilter](#)

Diagram: Wastewater treatment

Save image Print List Add concept Search Restore Quick guide Wastewater treatment Open

Update diagram Add box Edit Unlink Delete Save Close



<input type="checkbox"/> trickling filter	biological trickling filter; biological filter; biofilter; trickling biofilter;	biologisk filter;	Flows downward and is contacted with a ... A trickling filter consists of a fixed bed of rocks, gravel, slag, polyurethane foam, sphagnum peat moss, or plastic media over which sewage or other wastewater flows downward and is contacted with a ...	
<input type="checkbox"/> wastewater treatment plant	wastewater treatment works;	spildevandsanlæg; spildevandsrensningsanlæg;		
<input type="checkbox"/> wastewater treatment works	wastewater treatment plant;	spildevandsanlæg; spildevandsrensningsanlæg;		

<p>General</p> <p>New search</p> <p>Print</p> <p>Save to file</p>	<p>Select articles</p> <p>Select all</p> <p>Unselect all</p> <p>View selected</p>	<p>Action on selected</p> <p>Add to diagram</p> <p>Export selected</p> <p>Remove selected</p>	<p>Change display</p> <p>Sort alphabetically</p> <p>Group concepts</p> <p>Layout</p>
---	---	---	--

Authoring Memory

- Translation Memory database idea is used for source to source sentence comparison. A database with correctly structured sentences of a customer is compared (while writing) to the new text. Match rates indicate changes from previously used sentences.
 - Start Group – Mindreader
 - <http://www.star-group.net/ENU/mindreader/mindreader.html>
 - SDL – Author Assistant
 - <http://www.sdl.com/en/language-technology/products/global-authoring/sdl-authorassistant-client.asp>
 - Sajan - Authoring Memory Tool
 - <http://www.sajan.com/software/authoring-memory-tool.html>
 - Acrolinx – Intelligent Reuse Component

Authoring Memory function in Author-it Xtend

The screenshot shows the Author-it Xtend software interface. The main window displays a document titled "X1000 User Guide" with a table of contents on the left and a list of steps for "Assembling Your Teleportation Device" in the main area. The steps are:

1. Open the backpack by unzipping the zipper in the back (between shoulder straps).
2. Place the main module (Part A) into the backpack, with the label toward the outer zipper.
3. Snap the accessory mount (Part B) on top of the main module.
4. Pull the cords through the the color-coded openings that match the cord color.
5. To charge your teleporter, plug the blue cord into the wall uplink socket.
6. teleportation is the

A search window is open at the bottom right, displaying results for the term "teleportation". The search results are as follows:

Score	Text	Language
79 %	Teleportation is the fastest growing form of transportation	en
73 %	The X2001 is a portable "backpack-style" teleportation device.	en
72 %	This is the best teleportation device I have ever used!	en
72 %	ACME Teleportation Devices is the largest manufacturer and distributor of teleportation devices in the world.	en
70 %	The X5000 is the first teleportation device to meet ISO 13321 safety standard.	en

10 results found in 0,3 seconds

i-match: ITL's authoring support tool

Adobe FrameMaker (Strukturiert)

Datei Bearbeiten Element Format Darstellung Spezial Graphik Tabelle crossAuthorFM Fegster Hilfe

C:\Programme\Adobe\FrameMaker7.1\FrameMaker Beschreibung.fm

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Für Zaireiche Funktion zur Textverarbeitung
kombinird Frameamkr

i-match - sentences

Matches in [i]-write

1. Adobe FrameMaker 8 ist ein leistungsstarkes Werkzeug für technische Redakteure, das Funktionen zur Textverarbeitung sowie strukturiertes, XML-basiertes Authoring mit Publishing anhand von Schablonen kombiniert.

i-match - terms

Frameamkr x
FrameMaker (Software, proofed)

Textverarbeitung x
DTP (Software, proofed)

Standard File Formats

- Convergence of TMX and XLIFF?
 - TMX is used to exchange TM data
 - XLIFF is used as the format for translation, regardless of what format the file had originally.
 - Everything you can do with TMX, you can also do with XLIFF – so is there a need to keep both formats?
 - The TMX 2.0 specification was rejected so far, as it did not contain what users needed
- XLIFF is the file format of choice for open source tools
- XLIFF is also being used as a bilingual file format with some commercial translation tools (Heartsome, Swordfish, SDL Studio 2009, memoQ...)
- But XLIFF can contain so much more than just text to translate
 - History of changes, source of matches and match rates, alternative translations from comparison with different translation memories, for software files – size of GUI elements so that they could be shown visually during translation
 - But the tools that use XLIFF today are not really using all its potential

TMX

```
<tu tuid="1" datatype="Text" srclang="en-us">
  <tuv xml:lang="en-us">
    <seg>This is a test.</seg>
  </tuv>
  <tuv xml:lang="de">
    <seg>Dies ist ein Test.</seg>
  </tuv>
  <tuv xml:lang="ja">
    <seg>テストです。</seg>
  </tuv>
  <tuv xml:lang="zh-cn">
    <seg>这是试验。</seg>
  </tuv>
</tu>
```

XLIFF

```
<trans-unit id="n1">
  <source>This is a sentence.</source>
  <target xml:lang="de">Dies ist ein Satz.</target>

  <alt-trans match-quality="100%" tool="TM_System">
    <source>This is a sentence.</source>
    <target xml:lang="de">Das ist ein Satz.</target>
  </alt-trans>

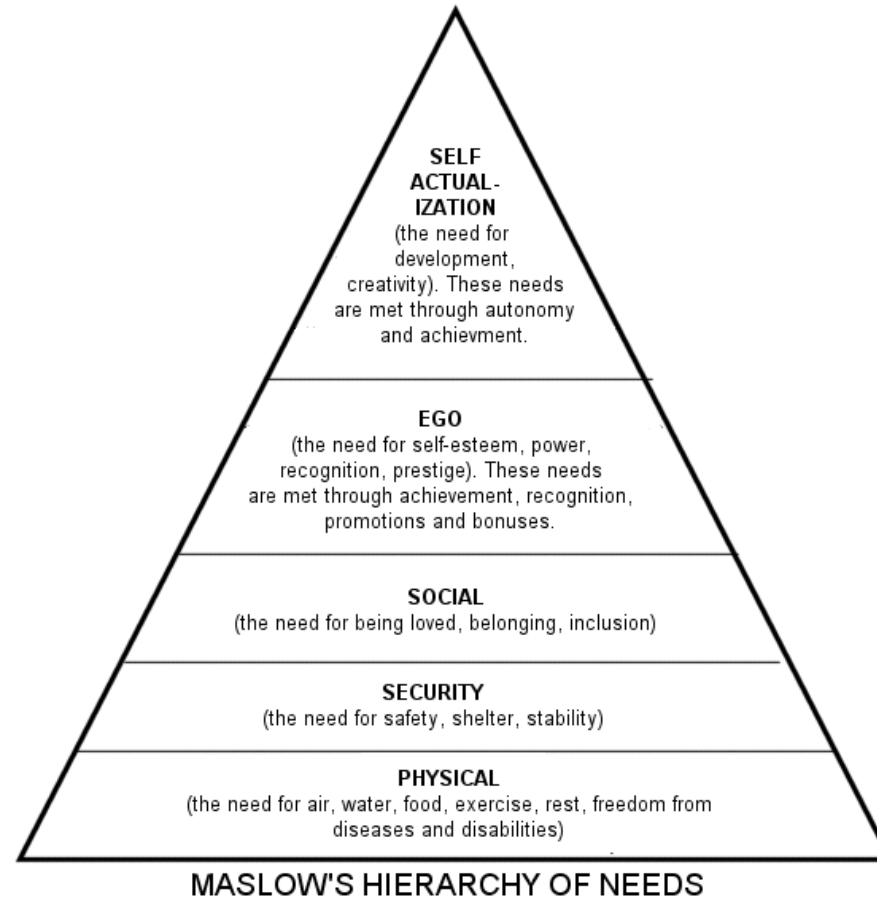
  <alt-trans match-quality="70%" tool="TM_System">
    <source>This is a short sentence.</source>
    <target xml:lang="de">Dies ist ein kurzer Satz.</target>
  </alt-trans>
</trans-unit>

<trans-unit id='1' resname='IDCANCEL'
  restype='button' coord='8;8;50;14'>
  <source xml:lang='en'>Cancel</source>
</trans-unit>
```

Crowdsourcing / community translation

- Projects like Facebook show, that it works, when there is a company or product
 - that has a community
 - that has a community that WANTS the content to be translated
 - where the individual user WANTS to contribute to the translated product (not for money but maybe prestige)
- LinkedIn on the other hand had a real problem because they wanted something for free
- It has happened for some time now for open-source software, freeware and shareware
- It was used for Harry Potter novels, Microsoft discussed terminology with its users...
- But, do you see that happening for the manual of a dishwasher or an electronic parts online shop?

Why Crowdsourcing Works?



Source:

<http://www.omafra.gov.on.ca/english/rural/facts/96-001.htm>

http://en.wikipedia.org/wiki/Maslow's_hierarchy_of_needs

Open source projects

- Mozilla Firefox
- Openoffice (Sun Microsystems)
- Ubuntu (<https://translations.launchpad.net/>)
- Highly structured organization: Contributors are selected, rated and monitored
- Very high quality (of both code and translation)

Crowdsourcing / community translation

- Tools for crowdsourcing are being made available
 - Facebook released its translation management and voting system to Facebook partner sites
 - CrowdSight for management of groups of translators and reviewers (add-on to GlobalSight, open-source translation tool by Welocalize)
 - Getting translators to sign up for a translation rather than asking individuals to do a certain translation with web portals like

Open Source

- Technology goes open source
 - Welocalize: GlobalSight translation management system
- Open source translation support tools
 - Okapi Framework
 - TM maintenance tool (Olifant)
 - Conversion to XLIFF (Rainbow)
- New developments
 - <http://www.opentms.de> (openTMS), open source translation memory system
- List of free and open source tools that support the translation process
<http://www.aspirationtech.org/papers/ott07/tools>

Links

- Article by Jost Zetsche on Google Translator Toolkit
 - <http://accurapid.com/journal/49google.htm>
- Comparative Study of statistical and example-based machine translation
 - <http://www.proz.com/translation-articles/articles/2483/1/Statistical-Machine-Translation-and-Example-based-Machine-Translation>
- MyGengo, Report on the state of the translation industry 2009
http://workfortranslators.files.wordpress.com/2009/10/mygengo_state_of_translation_industry_2009.pdf



Thank you

Angelika Zerfaß
zerfass@zaac.de