MyMemory

Creating the world’s largest Translation Memory

Marco Trombetti
Translated
MT Marathon
Trento, Sep 2011
http://mymemory.translated.net
Language Service Provider
Human translation in 80 languages, 15,000 customers, 170,000 translations delivered.

Core Technology
A fully automated translation workflow. Customers and LSPs can access 45,000 translators through a web form or an API.
2 Challenges
1. Help Translators

2. Cut out the middle man
How?
Translation Memories fail!

Prob(10-word-sentence) is too small
Machine Translation Market

- Human Translation: 99%
- Machine Translation: 1%

Source: Allied Business Intelligence
# My wish-list for the perfect TM Platform

<table>
<thead>
<tr>
<th>Feature</th>
<th>MyMemory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search across all memories, and still be FAST</td>
<td>✔️</td>
</tr>
<tr>
<td>Web 2.0, Wiki-like contributions</td>
<td>✔️</td>
</tr>
<tr>
<td>One-click to download a TM</td>
<td>✔️</td>
</tr>
<tr>
<td>Adaptive Matching, A priori probability</td>
<td>✔️</td>
</tr>
<tr>
<td>Propagate changes in real-time</td>
<td>✔️</td>
</tr>
<tr>
<td>Fully integrated Statistical MT</td>
<td>✔️</td>
</tr>
<tr>
<td>Open standard and CAT support</td>
<td>✔️</td>
</tr>
</tbody>
</table>
The Solution

Free, Collaborative Web Translation Memory
Barriers to contribution

Privacy and IP

Time vs Value
Will people contribute?

Yes!

100 words per second during the last year

5 billion words
Contribution Trend

Million of segments

0
50
100
150
200
250
300
350
400
450
500
Sources

- Web
- Translators
- Customers
How many users yesterday?

74,399 Unique visitors

960,000 Queries
Benefits

LSPs & Clients
✓ More control
✓ Free contribution from pros & the crowd
✓ Quick reference and QA

Translators
✓ Quick, relevant reference
✓ Improved organization and usability
✓ Free memories
Technology Challenges

- Scalable Real Time Indexing
- Adaptive Ranking
- Cost Effective Scalability
Generating the interest

Pre-Populate

Make contributing easy

Make it free
Protecting IP

- Private Memories
- Hide Proper Nouns
- One-click to report abuse
- Robots.txt
Quality

- Feedback from the crowd
- QA Correlation
- Automatic QA
Is 5 billion words enough?

No.

The language industry produces tens of billions of words per year.
Every CAT tool connected
Brave
Live Demo

http://mymemory.translated.net