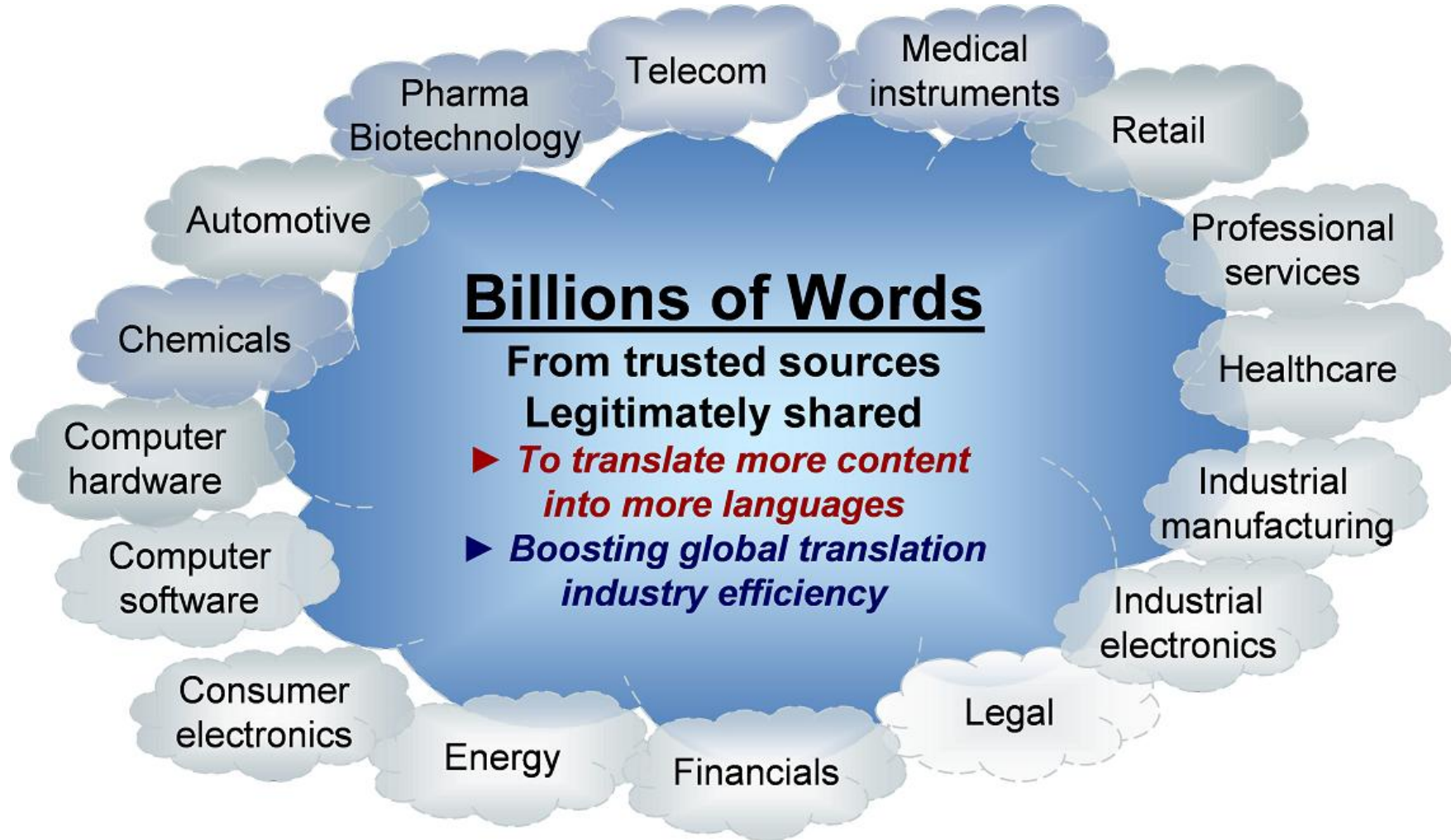


Implement an Industry Vision

One global platform open to all stakeholders in the translation industry



TDA Services

- **TM Sharing Services** – *for members only*
 - Upload and download translation memories

- **Search** – *free & open to the public*
 - Look-up translations of terms and phrases

Benefits of Search

- Increase quality and speed of translation
- Resolve QA bottlenecks
- Resource for support and engineering
- Streamline industry terminology
- Translators training and research

Benefits of TM Sharing

- Advanced leveraging: 35% to 50%
- Improved performance of machine translation: 50% jump in BLEU score
- Springboard for new value-added services

How People Use TDA Today

Start with uploading your TMs

	Action	Result	Saving
TERMINOLOGY	Upload your TMs. If you are concerned about sharing TMs for other members to download, you can tick the box "For Search Only". Ask your translators and reviewers to use TAUS Search to look up terms and phrases.	It seems so obvious, but most people can't look up terms and phrases in the whole corpus of their TMs. TAUS Search lets everyone find translations in all of your uploaded TMs, and they may opt to search across industry. This will help to solve translation and review bottlenecks, saving time, increasing quality and consistency.	5%-10%
TRANSLATION MEMORY	Select TMs for downloading by industry or data owner while checking the volume counter. The success of additional leveraging is dependent on finding sufficient proximate language data. You can import the TMX files in your regular translation editor and start leveraging translations.	If you get less than 5% matches from the TMX files you have downloaded from TDA, you may want to try another translation tool. Leveraging translations from large TM corpora is different than the traditional project-based TM approach. Phrase-based leveraging, supported by statistical routines and linguistic intelligence in a corpus-TM environment can generate 10% to 50% or more high-fuzzy matches.	10%-50%
MACHINE TRANSLATION	Select TMs for downloading by language pair, data owner, industry and/or content type. You can use the TMX files for the training of MT engines. The TMX tags usually need to be removed. You can use the TM files for the training of the "translation models" and also use the target side for the training of the "language model". The size of the corpus to be used for training depends on the engine and other factors. Hybrid rule-based models usually require smaller volumes than statistical engines.	The success of MT training is measured in metrics such as BLEU score. Pilot projects have shown significant increases in quality of up to 50% as a result of using much larger collections of data from TDA. Good quality MT output can double or quadruple the translation/post-editing productivity, or allow publishers to provide real-time fully automatic translation of for instance support content.	50%

Benefits from terminology (TAUS Search) can be obtained easily and quickly. The benefits from TDA for translation memory and machine translation require planning and investment of time and resources. Twenty out of the sixty current members seem to be making these investments at the moment, whether directly or via their language service providers.

Ideas for New Features & Services

	New service or feature	Benefits
TERMINOLOGY	Multi-word translation. Currently we compute translations for single words only. Extending this computed translation to include phrases.	Better translation quality and saving more time and cost.
	Synonym search. Allow TAUS Search to automatically find related terms and their translations in context.	Better translation quality and saving more time and cost.
	Matrix search. Allow to search across all language pairs (instead of primarily from and into English).	Make TAUS Search beneficial for more users and more languages.
TM	Tool compliance. Currently all TMs are stored in a neutral TMX format. This feature allows users to also store TMs in the tool compliant format, optimizing the leveraging within the same tool.	TDA can be used for all TM sharing by virtual translation teams without using any leveraging.
	Translation Matching. This feature allows users to upload new documents and retrieve all matches from the entire TDA repository in a TMX format.	Easy way to retrieve all matches from the entire database. Bonus on matches.
TM & MT	TM Cleaning. A statistical tool that filters out suspicious translation units.	Eliminate bad quality translations.
	Matrix TM. Allow users to extract TMs from TDA in all language pairs (as long as data owner and product line correspond).	Allow TM leveraging in new languages.
	Matching scores. A statistical tool that allows users to identify the best matching data for a particular job and to zoom in or out depending on the volume or accuracy requirements.	Ideal for optimizing data selection.
MT	MT Trainer. Allows users to upload TMs and request new engines to be trained through TDA users based on TDA data sets.	Access and compare engines for all languages and niches.
	Genre identification. Statistical tool that identifies content types, helping users to select data of the same genre for MT training.	Ideal for optimizing data selection.

Private: members may limit sharing of TMs to their own selection of registered users ('private vaults').

Integration: API's for all services will be available to everyone.



Development Priorities

Please indicate below the priorities that you would like to give to the planned features described in the TDA Roadmap 2010 document.

	Highest priority	High priority	Lower priority	Lowest priority	Average score	Priority
Tool compliance	10.5% (2)	57.9% (11)	26.3% (5)	5.3% (1)	2,26	6
Matching Scores	25.0% (5)	50.0% (10)	20.0% (4)	5.0% (1)	2,05	3
TM Cleaning	45.0% (9)	35.0% (7)	15.0% (3)	5.0% (1)	1,80	1
Translation Matching	42.1% (8)	36.8% (7)	15.8% (3)	5.3% (1)	1,84	2
API Translation Matching	25.0% (5)	50.0% (10)	20.0% (4)	5.0% (1)	2,05	3
Search Plug-in	40.0% (8)	15.0% (3)	45.0% (9)	0.0% (0)	2,05	3
MT Trainer & Evaluator	40.0% (8)	15.0% (3)	30.0% (6)	15.0% (3)	2,20	5
API MT Trainer	30.0% (6)	25.0% (5)	30.0% (6)	15.0% (3)	2,30	7
Synonym Search	20.0% (4)	35.0% (7)	30.0% (6)	15.0% (3)	2,40	8
Multi-word Translation	20.0% (4)	55.0% (11)	15.0% (3)	10.0% (2)	2,15	4
Matrix Search	10.0% (2)	35.0% (7)	45.0% (9)	10.0% (2)	2,55	9
Matrix TM Repository	10.0% (2)	40.0% (8)	35.0% (7)	15.0% (3)	2,55	9

Strategic Actions

- Adjustment annual fees in line with size of operation and realizable benefits. Moderate increase for 'large members'
- Make API's publicly available
- Open sourcing all TDA software components
- Open for sponsoring and funding
- Open to TM Sharing (For Search Only) for 'fair use'
- Partner Agreements for data & member acquisition
- Development priorities
 - Translation Matching (sponsored)
 - TM Cleaning
 - Matching Scores

How can you contribute and participate

- Use Search – API integration
- Share translation memories – API integration
- Join TDA as a new member