

Achim Ruopp
Digital Silk Road LLC
achim@digitalsilkroad.net

BIGTM.NET

GAINING A HEAD START ON TRANSLATION PROJECTS

A New Translation Project

- “The Bundle Breaker® is a sturdy machine with moving parts driven by electric motors, pneumatics, and hydraulics. These parts move automatically when the machine is turned on.”
- “The downstream platen of the compression bridge may move side-to-side unexpectedly and strike personnel in its path.”



A New Translation Project Bi-Lingual Terminology

- ⦿ Criteria
 - In domain
 - Correct
 - Current
 - In context
- ⦿ Sources
 - Translation memory from previous projects
 - Domain dictionaries
 - ... BigTM.net

BigTM.net Custom Translation Search

- ⦿ What is it?
 - Custom Translation Search Engine
 - Input: project source text
 - Searches the web for **similar** bi-lingual pages
 - Indexes discovered bi-lingual pages
 - Provides search UI
- ⦿ Current Language Pairs
 - English - French / Italian / German / Spanish

BigTM.net Project Page

Projects

ID	Created	Name	Description	Source	Target	Step		Term Cloud
1	2009-07-02 22:02:05	Bundle	Bundle Breaker Manual	English	Spanish	7/7	Search	air area automatic bolt breaker bridge bundle bundles button compression conveyor display downstream drawbridge electrical energy fault hazardous help hydraulic information lockout machine maintenance manual operation operator page pin platen press pressure procedure safety screen stop touch —
2	2009-07-09 15:23:35	Beispiel	Beispiel.com website content	English	German	0/7		
3	2009-07-09 15:24:44	Bike	Motorcycle Repair Manual	English	Italian	0/7		

Done

BigTM.net Search Results

Search Result for "hydraulic"

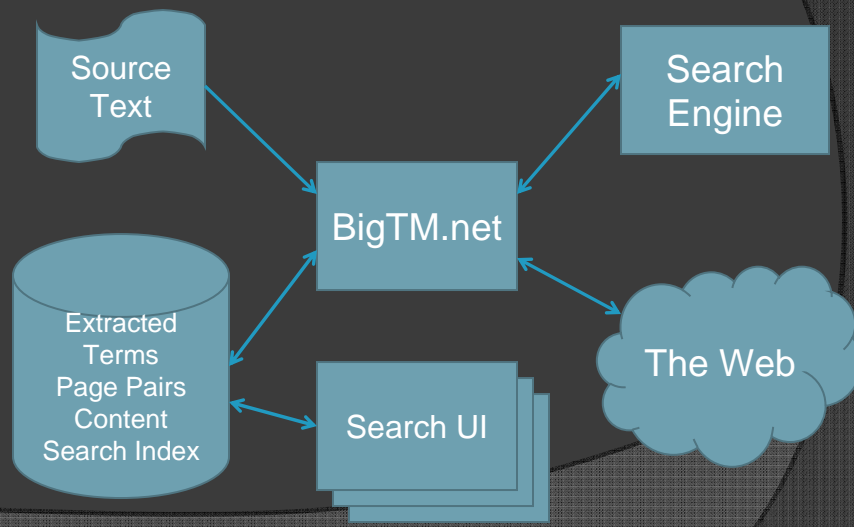
http://www.cascorp.com/eme/en/hydraulicbinretainer	http://www.cascorp.com/eme/es/hydraulicbinretainer
Cascade Corporation Hydraulic Bin Retainer	Cascade Corporation Rotonodor hidráulico de contenedores
Watch a video clip of the Hydraulic Bin Retainer. Choose:	Ver clip de vídeo del retenedor hidráulico de contenedores. Elija:
The Hydraulic Bin Retainer accommodates plastic and wood bins used in the agricultural markets such as handling crates, apples, pears, potatoes, nuts and seeds.	El retenedor hidráulico de contenedores permite utilizar los contenedores plásticos y de madera que se utilizan en los mercados agrícolas para manipular uva, manzanas, peras, patatas, nueces, semillas, etc...
The Hydraulic Bin Retainer is also popular with the industrial sector handling steel bins used to transport scrap metal and paper waste.	El retenedor hidráulico de contenedores también es muy utilizado en el sector industrial que emplea contenedores metálicos para transportar desechos de metal y de papel.
http://www.cd3wd.com/cd3wd_40/vita/vthbook/vp/vthbook.htm	http://www.cd3wd.com/cd3wd_40/vita/vthbook/es/vthbook.htm
A hydraulic ram is a self-powered pump that uses the energy of falling water to	Un camero hidráulico es una bomba mismo-impulsada que usa la energía de caerse el agua a
the use of commercial hydraulic rams, which are available in some countries. Plans	el uso de cameros hidráulicos comerciales que están disponible en algunos países. Los Plance
for building your own hydraulic ram are also available from VITA and elsewhere.	por construir su propio camero hidráulico también está disponible de VITA y en otra parte.
Use of the Hydraulic Ram	El uso del Camero Hidráulico

Done

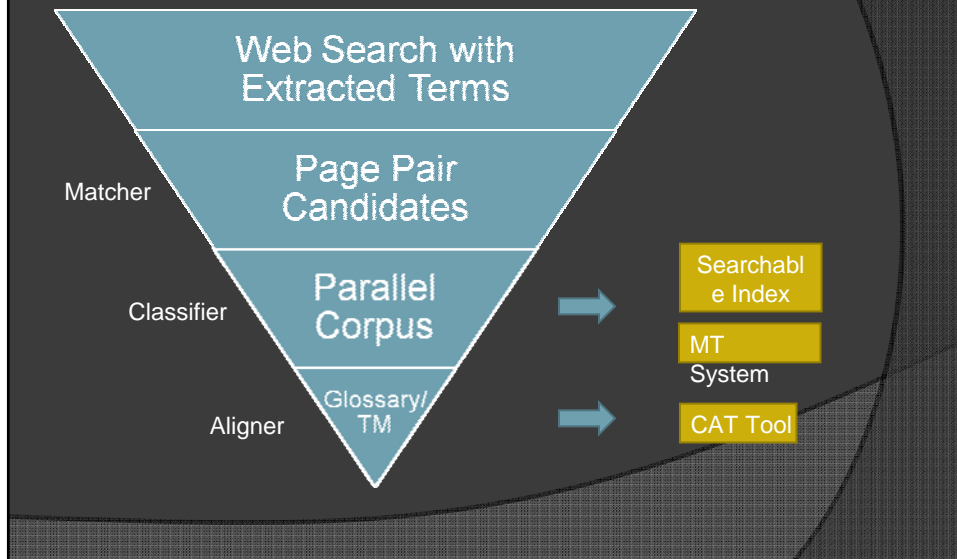
Terminology Criteria

- In domain
 - BigTM.net searches the web for similar parallel pages
 - Search UI results are ranked according to similarity with source text
- Correct
 - Translator can choose from discovered variants
 - Relies on web data
- Current
 - Web search turns up new translations
- In context
 - Search results always displayed in context

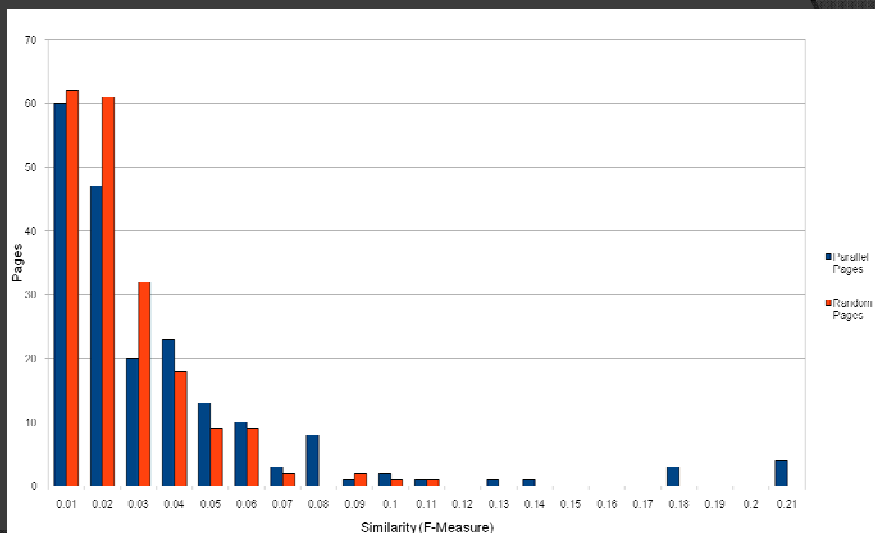
BigTM.net Architecture



BigTM.net Data Flow



Similarity with Source Text



Integration with Translation Tools

- ◉ Tools with web search integration
- ◉ Dictionary download
 - CSV format with probabilities
- ◉ Translation memory download
 - Standard TMX format
 - Best used with sub-segment matching

Training of domain-specific statistical MT systems

- ◉ Supplement general domain corpus with domain-specific corpus downloaded from BigTM.net
- ◉ KDE documentation prototype English-German

Corpus	MT System	BLEU
~ 600k sentences Europarl	Moses	12.81
~ 400k sentences Europarl ~ 200k sentences from BigTM	Moses	16.47 (+3.66)
n/a	Google	18.29 (+5.48)

Benefits Summary

- Provides a searchable index for translations of similar texts
- Rapidly prototypes terminology for review by the customer
- Provides a core translation memory as a starting point
- Allows the training of domain-specific machine translation systems

References

- Website: <http://www.bigtm.net/>
- Email: bigtm@digitalsilkroad.net
- Email your source text to get added to the beta program