

Cross Lingual Retrieval on Wikipedia



Stéphane Clinchant
stephane.clinchant@xrce.xerox.com

Xerox Research Center Europe
Laboratoire d'Informatique de Grenoble

Context

- Outcome of european project: **SMART:**
 - <http://smart-project.eu>
 - Objective: build a bridge between
 - Machine Learning
 - Machine Translation, Cross-Lingual Retrieval
- Goal of the Demo:
 - Highlight new techniques from SMART:
 - Mainly *Sinuhe* (SMT)
 - Demonstrate that they can scale up

A Wikipedia Demo

- Fact:
 - Lots of languages have few contributions in wikipedia, and the link in other languages is often not reliable
- Original Scenario:
 - Provide slovenian users access to the french wikipedia
 - Any pair of language could be added
 - Ex: english to french.


Demo Instructions

- Go to <http://cosco-demo.hiit.fi/smart>
- Type your query:
- Click on the title of a result to start the translation (from french to english)

You can also choose the machine translation engine:

- Moses (state of the art SMT)
- Sinuhe (new SMT system from SMART)

Step 1



S M A R T
Statistical Multilingual Analysis
for Retrieval and Translation

French Wikipedia '06	▼	search collection
English	▼	query language
English (sinuhe_giga)	▼	target language
UH	▼	CLIR search engine

machine translation



Type your query here



SMART

Statistical Multilingual Analysis
for Retrieval and Translation

French Wikipedia '06 search collection

English query language

English (sinuhe_giga) target language

UH CLIR search engine

machine translation

Search

1-10 results (took 3.364 seconds)

Machine translation

Traduction automatique

Machine translation is in the strict sense of the fully translated into a text through one or more computer programs, without a translator human has not to into comparable to that of translators human is still of utopia.

La traduction automatique désigne au sens strict le fait de traduire entièrement un texte grâce à un ou plusieurs programmes informatiques, sans qu'un traducteur humain. Les traductions d'une qualité comparable à celle de traducteurs humains relève encore de l'utopie.

Translation

Traduction

Translation is to interpret the meaning of a text in a language (source language, or language of departure), and produce a text with a meaning and a effect e arrival).

La traduction est le fait d'interpréter le sens d'un texte dans une langue (langue source, ou langue de départ), et de produire un texte ayant un sens et un effet équivalent.

Translation memory

Mémoire de traduction

Definition. A translation memory is a database containing the segments of text as well as the equivalent of these segments in another language. It allows to :

Définition. Une mémoire de traduction est une banque de données contenant des segments de texte ainsi que l'équivalent de ces segments dans une autre langue. Elle permet :

English translation into French

Traduction de l'anglais vers le français

Prior to read. The translator must not only know the source language (in English) but also, and especially the target language (by French). It is absolutely nec

Avant de lire. Le traducteur doit non seulement bien connaître la langue source (ici l'anglais) mais également, et surtout, la langue cible (ici le français). Il est absolu

Translation of languages distributed

Traduction de langues distribuée

The project Translation of Distributed languages (English: Distributed Language Translation, in Esperanto Distribuita Lingvo-Tradukado, abbr DLT) was a Trans airline Dutch Fokker and directed by society Bos in Utrecht.

Le projet Traduction de langues distribuée (en anglais : Distributed Language Translation, en espéranto Distribuita Lingvo-Tradukado, en abrégé DLT) était un projet de compagnie d'aviation néerlandaise Fokker et réalisé par la société BSO à Utrecht.



S M A R T

Statistical Multilingual Analysis
for Retrieval and Translation

French Wikipedia '06 search collection

English query language

English (sinuhe_giga) target language

UH CLIR search engine

machine translation

Search

1-10 results (took 3.364 seconds)

Machine translation

Traduction automatique

Machine translation is in the strict sense of the word not comparable to that of translators human is still

La traduction automatique désigne au sens strict les traductions d'une qualité comparable à celle de tra

Translator human has not to inte

ques, sans qu'un traducteur h

Click to launch translation
of the page

Translation

Traduction

Translation is to interpret the meaning of a text (in the arrival).

La traduction est le fait d'interpréter le sens d'un t

with a meaning and a effect e

tant un sens et un effet équiv

Translation memory

Mémoire de traduction

Definition. A translation memory is a database conta

Définition. Une mémoire de traduction est une banque

in another language. It allows to :

segments dans une autre langue. E

Original Page in French

English translation into French

Traduction de l'anglais vers le français

Prior to read. The translator must not only know the

Avant de lire. Le traducteur doit non seulement bien c

re (by French). It is absolutely nec

cible (ici le français). Il est absolur

Translation of languages distributed

Traduction de langues distribuée

The project Translation of Distributed languages (English: Distributed Language Translation, in Esperanto Distribuita Lingvo-Tradukado, abbr DLT) was a Trans airline Dutch Fokker and directed by society Bos in Utrecht.

Le projet Traduction de langues distribuée (en anglais : Distributed Language Translation, en espéranto Distribuita Lingvo-Tradukado, en abrégé DLT) était un projet c compagnie d'aviation néerlandaise Fokker et réalisé par la société BSO à Utrecht.

What you can do with the Demo

- Assess quality of translation:
 - Can you get most of the information ?
 - Is the translation readable ?
- Compare Sinuhe vs Moses:
 - wrt Speed
 - wrt Quality of translation
- If you know some french, you can compare the original and the translation

Conclusion

- Demo at <http://cosco-demo.hiit.fi/smart>
- Come for a live Demo or any questions !