WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Patent Machine Translation (Handling large data with Moses)

Bruno Pouliquen (Bruno.Pouliquen@wipo.int)
with the help of Marcin Junczys-Dowmunt

What this talk is about?

- Introduction
- Patent translation, what is specific about it?
- Our tool: Tapta
- Big model management with Moses
- Language specificities
- Our tool installed in various places
- User interfaces
- Quality / user acceptance
- Conclusion

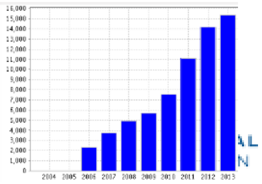
Introduction: Basic facts about WIPO

World Intellectual Property Organization

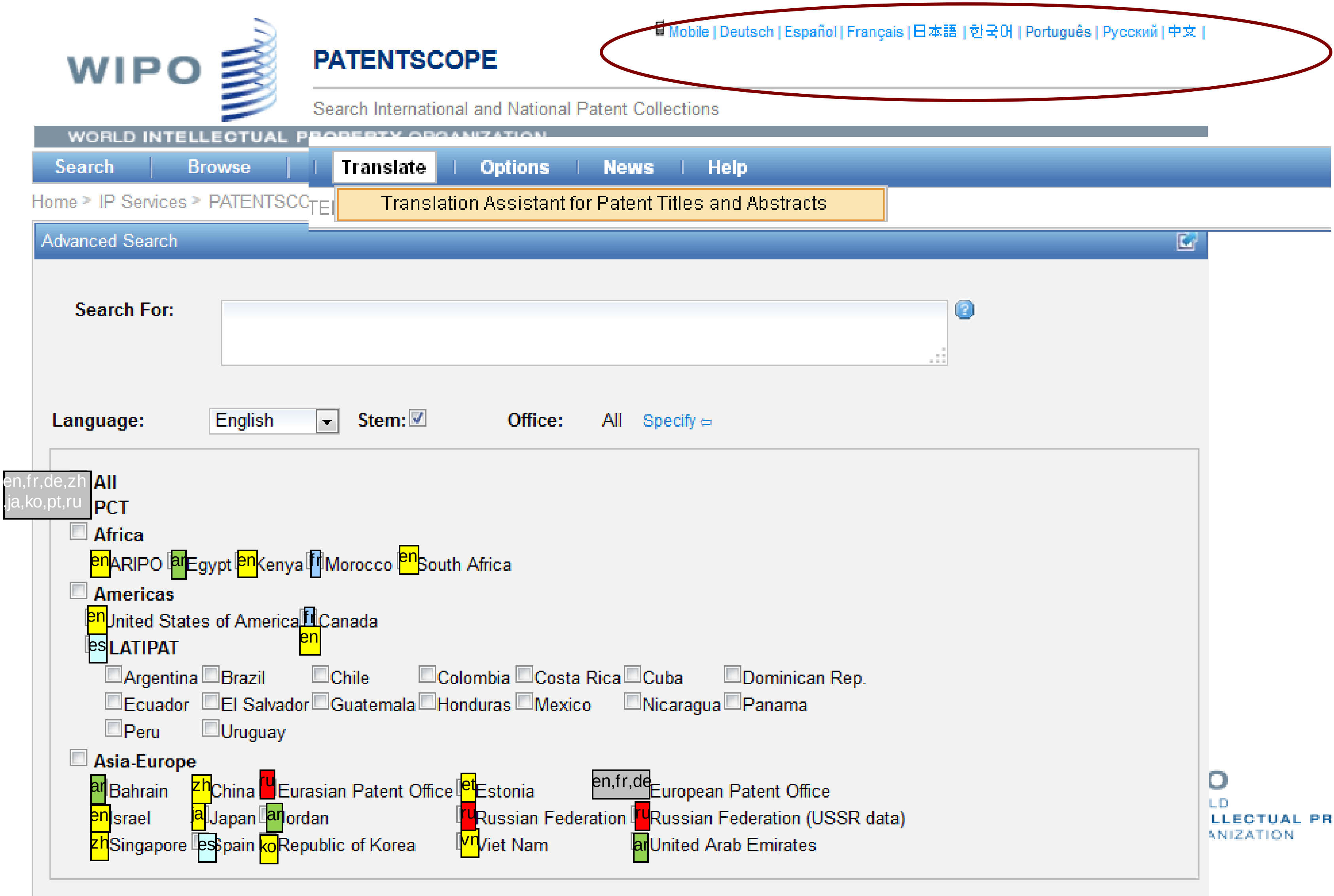
Mission : promote the protection of intellectual property rights worldwide and extend the benefits of the international intellectual property system to all member states

Status : a specialized agency of the UN
Member states : 184
Observers : 250+
Staff : 950 from 101 countries

Translation of Patents:
PCT : translation of titles/abstracts and International Search reports (40 million words/year to EN and FR, 90% outsourced)



Patentscope: patent search engine



Mobile | Deutsch | Español | Français | 日本語 | 한국어 | Português | Русский | 中文 |

WIPO

WORLD INTELLECTUAL PROPERTY ORGANIZATION

PATENTSCOPE

Search International and National Patent Collections

Search | Browse | Translate | Options | News | Help

Translation Assistant for Patent Titles and Abstracts

Home > IP Services > PATENTSCOPE

Advanced Search

Search For:

Language: English **Stem:** ☑ **Office:** All Specify

en,fr,de,zh,ja,ko,pt,ru **All**

PCT

□ **Africa**

en ARIPO | ar Egypt | en Kenya | fr Morocco | en South Africa

□ **Americas**

en United States of America | fr en Canada

es LATIPAT

□ Argentina □ Brazil □ Chile □ Colombia □ Costa Rica □ Cuba □ Dominican Rep.

□ Ecuador □ El Salvador □ Guatemala □ Honduras □ Mexico □ Nicaragua □ Panama

□ Peru □ Uruguay

□ **Asia-Europe**

ar Bahrain | zh China | ru Eurasian Patent Office | et Estonia | en,fr,de European Patent Office

en Israel | ja Japan | ar Jordan | ru Russian Federation | ru Russian Federation (USSR data)

zh Singapore | es Spain | ko Republic of Korea | vn Viet Nam | ar United Arab Emirates

Patent translation, what is specific about it?

- A patent application is made up of a title, an abstract, a description and claims
  - Title: 8 words

    Writing instrument with clip and pencil
  - Abstract: 114 words
  - Description: 6'428 words!

    The clipboard pencil holder is a device that can be easily attached to the clip of a conventional clipboard for the purpose of providing easy and convenient storage and retrieval of a pencil or other similarly-sized writing instrument. The holder holds the pencil horizontally against, and parallel to the clipboard clip while substantially avoiding interference with the clipboard clip's ability to grip a paper pad which, for example, is positioned between the clip and the board of the clipboard.
  - Claims: 726 words

    Claim 1: A writing instrument, which consists of a pencil and clip for holding the apparatus in a product, said clip being attached to said pencil.
    Claim 2: The writing instrument of claim 1 wherein said clip may be fixed to said pencil
    …
- Meta information: office, classification (IPC), original language, filing date, publication date, inventor etc…
- Specific language, scientific terms, almost no repetitive text (Unknown phrase: "I am", almost no proper names…)
- Usually only the title and the abstract are translated

WIPO SMT tool: TAPTA

Requirements

- Fully automatic
  - preparation of data
  - training
  - evaluating/mert/binarization/etc.
- Domain aware
- Fast translations (on the fly)
- Free to use (open source + in-house development)
- Confidentiality
- Various User interfaces
- First goal: assimilation, online translation of patent applications on our search engine
- Additional goal: dissemination, integration in CAT tool, “translation accelerator”

TAPTA: First version - 2011

- Patent SMT: title+abstracts
- 180 M words (en-fr), 8M words
- Called TAPTA ("Translation Assistant for Patent Title and Abstract")
- Domain-aware: 32 domains encoded as factors in Moses
- Pouliquen et al. EAMT 2011

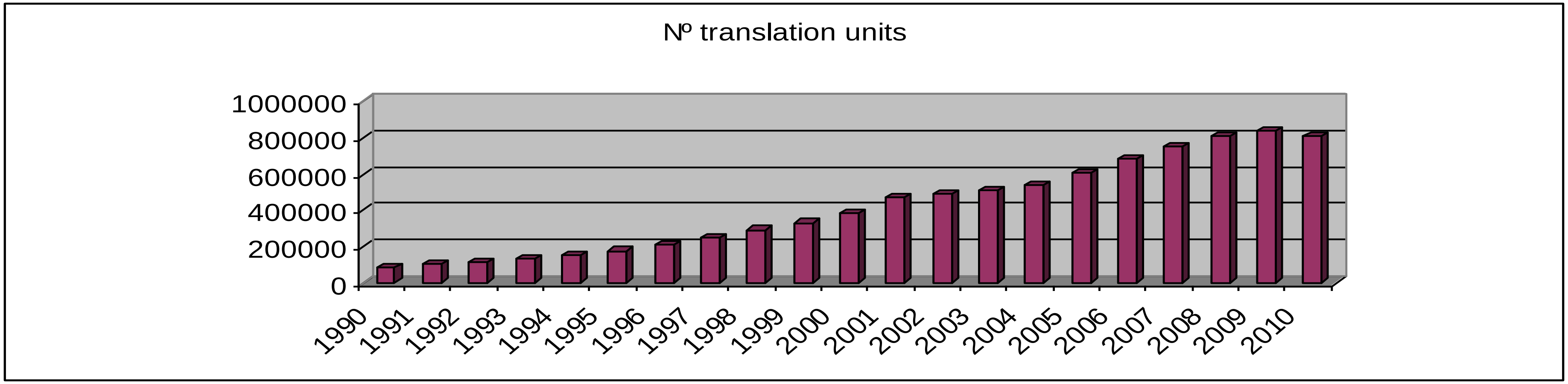
COPPA: **Corpus Of Parallel Patent Applications**

All English-French PCT application title and abstract (1990-2010)

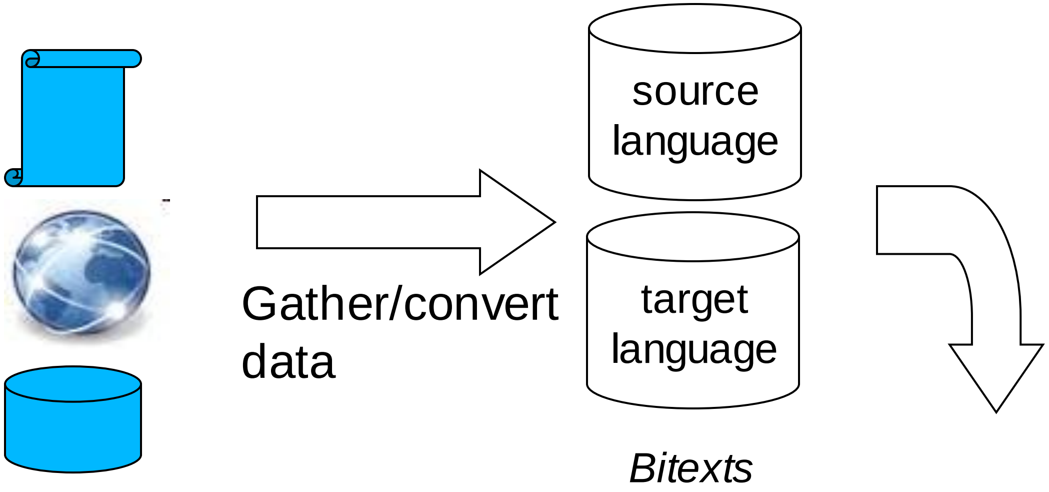
Free for research

180 Million words in TMX format (8.7 Million translation units)

http://www.wipo.int/patentscope/en/data/products.html#coppa



Tapta framework



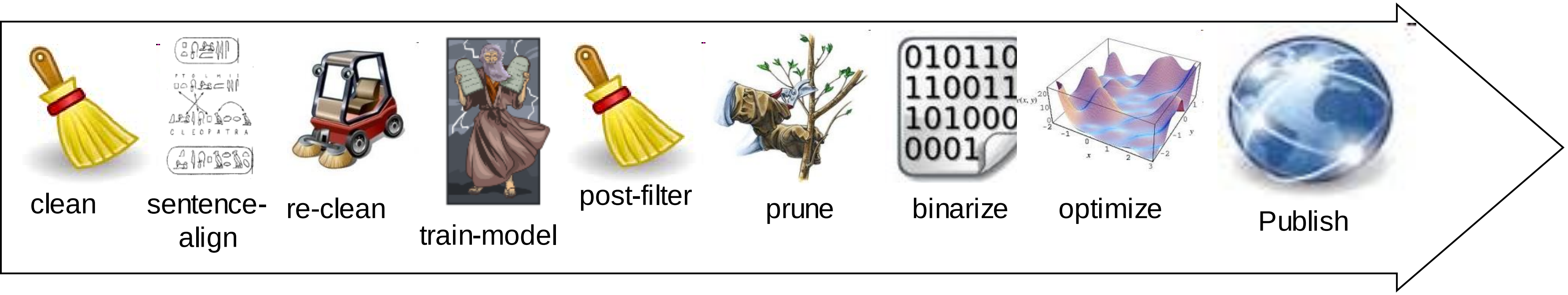
Gather/convert data

source language

target language

Bitexts

Our system prepares the data for Moses, applies some post-processing (filter, pruning, binarization, optimization…) and offers various interfaces to translate



clean

sentence-align

re-clean

train-model

post-filter

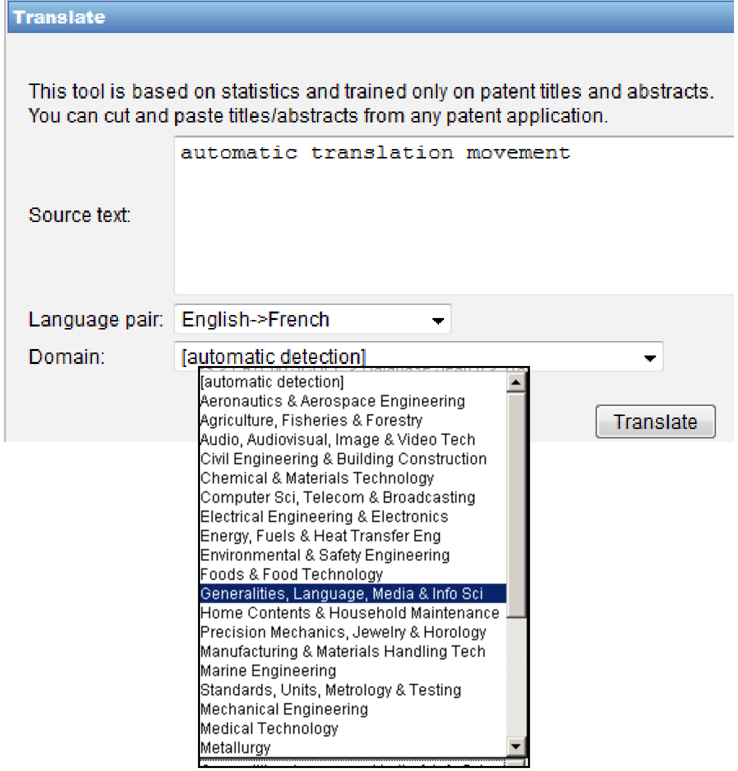
prune

binarize

optimize

Publish

TAPTA:



Translate

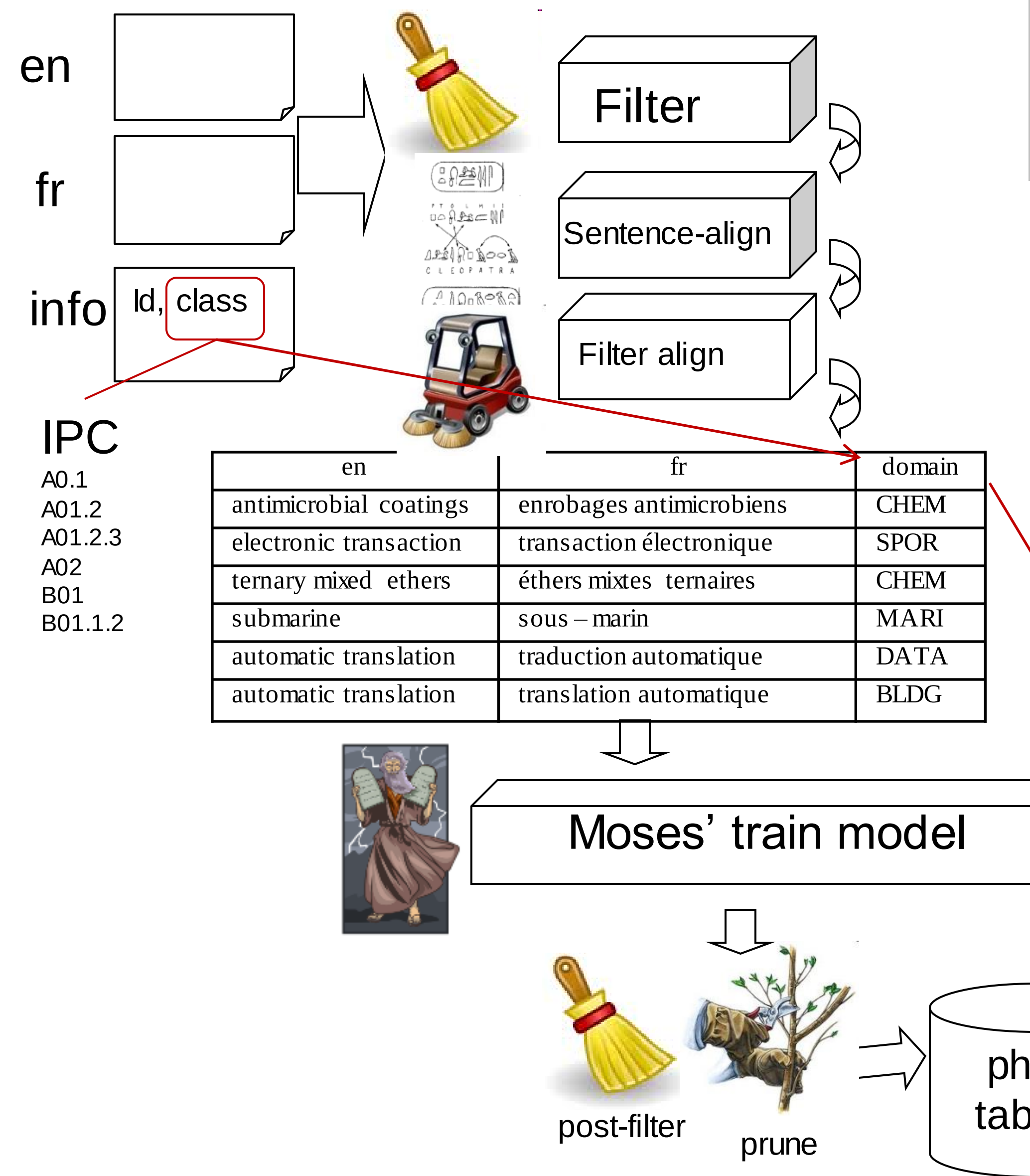
This tool is based on statistics and trained only on patent titles and abstracts.
You can cut and paste titles/abstracts from any patent application.

Source text: automatic translation movement

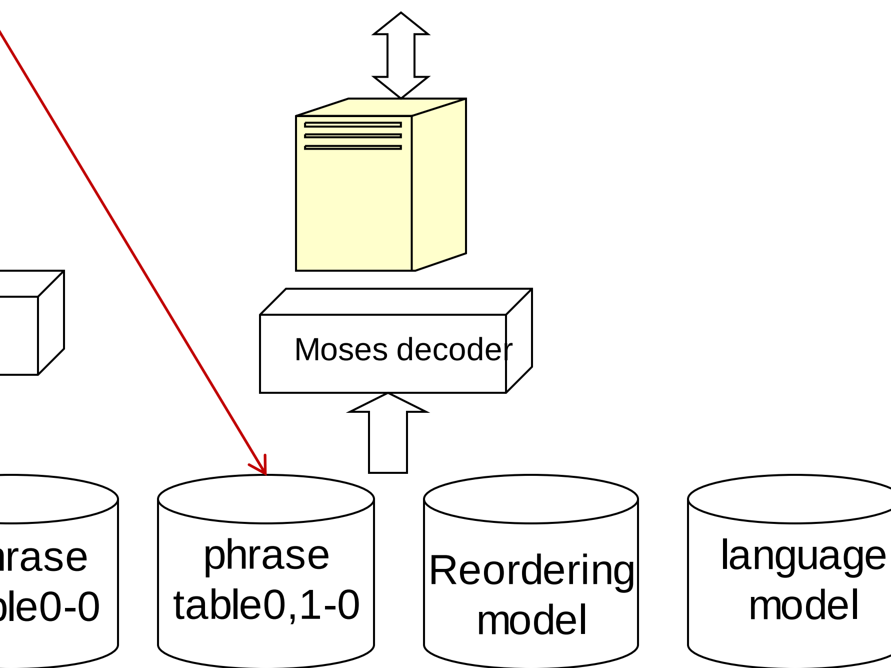
Language pair: English->French

Domain: [automatic detection]

- [automatic detection]
- Aeronautics & Aerospace Engineering
- Agriculture, Fisheries & Forestry
- Audio, Audiovisual, Image & Video Tech
- Civil Engineering & Building Construction
- Chemical & Materials Technology
- Computer Sci, Telecom & Broadcasting
- Electrical Engineering & Electronics
- Energy, Fuels & Heat Transfer Eng
- Environmental & Safety Engineering
- Foods & Food Technology
- Generalities, Language, Media & Info Sci
- Home Contents & Household Maintenance
- Precision Mechanics, Jewelry & Horology
- Manufacturing & Materials Handling Tech
- Marine Engineering
- Standards, Units, Metrology & Testing
- Mechanical Engineering
- Medical Technology
- Metallurgy



en	fr	domain
antimicrobial coatings	enrobages antimicrobiens	CHEM
electronic transaction	transaction électronique	SPOR
ternary mixed ethers	éthers mixtes ternaires	CHEM
submarine	sous – marin	MARI
automatic translation	traduction automatique	DATA
automatic translation	translation automatique	BLDG

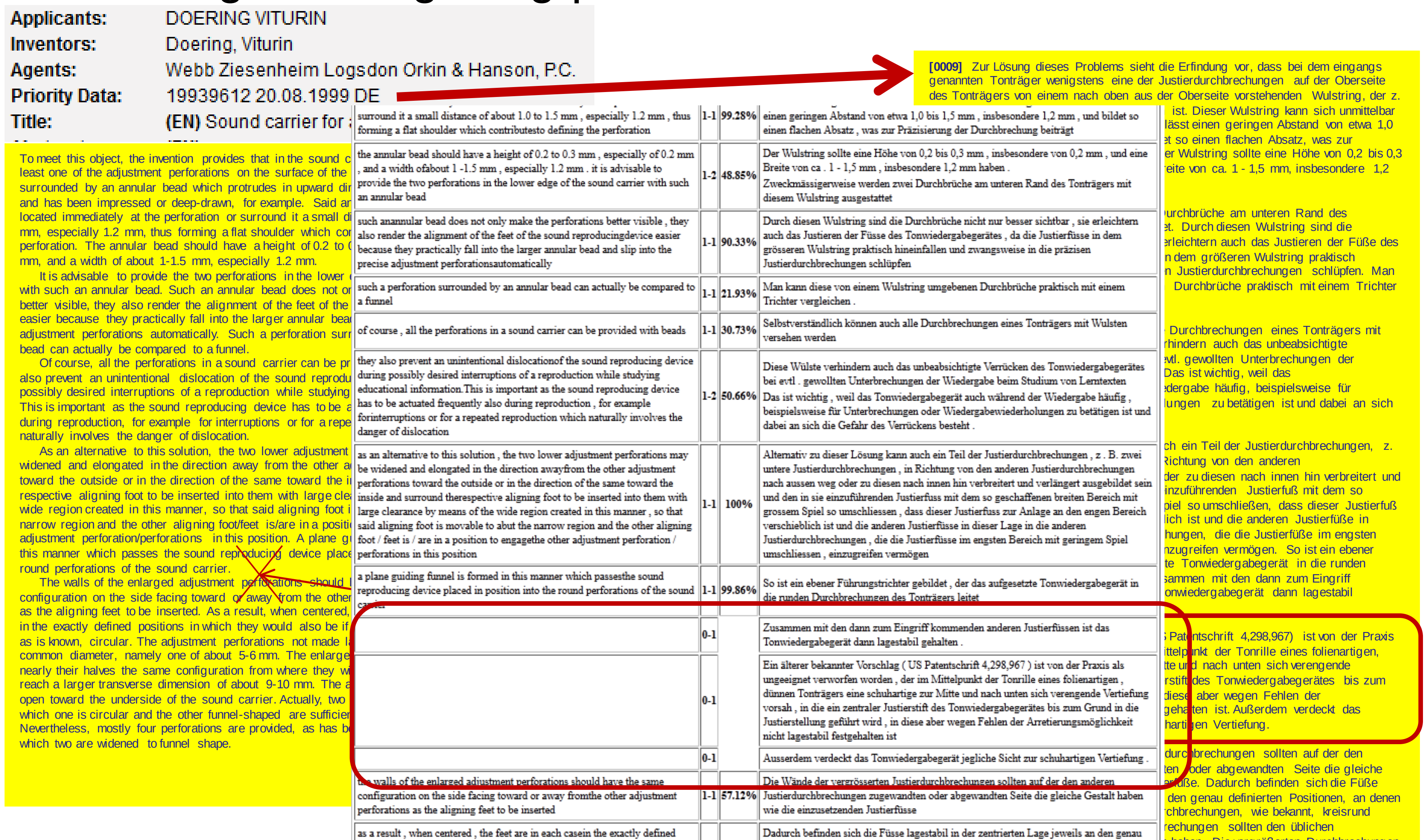


TAPTA: more than titles & abstracts

- Many offices, many languages (en, ja, fr, zh, de, ru, ko, es, pt…)
- Usually only the title and abstract are translated…
- But:
  - European Patent Office: translations of claims in English, French and German
  - One invention can be patented in more than one country (in various languages)…
  - Huge amount of data ~ Millions of patent applications

Description alignment

Challenges with getting parallel sentences



Applicants:	DOERING VITURIN
Inventors:	Doering, Viturin
Agents:	Webb Ziesenheim Logsdon Orkin & Hanson, P.C.
Priority Data:	19939612 20.08.1999 DE
Title:	(EN) Sound carrier for a

To meet this object, the invention provides that in the sound c
least one of the adjustment perforations on the surface of the
surrounded by an annular bead which protrudes in upward dir
and has been impressed or deep-drawn, for example. Said an
located immediately at the perforation or surround it a small di
mm, especially 1.2 mm, thus forming a flat shoulder which con
perforation. The annular bead should have a height of 0.2 to 0
mm, and a width of about 1-1.5 mm, especially 1.2 mm.

It is advisable to provide the two perforations in the lower e
with such an annular bead. Such an annular bead does not on
better visible, they also render the alignment of the feet of the
easier because they practically fall into the larger annular bead
adjustment perforations automatically. Such a perforation surr
bead can actually be compared to a funnel.

Of course, all the perforations in a sound carrier can be pr
also prevent an unintentional dislocation of the sound reprodu
possibly desired interruptions of a reproduction while studying
This is important as the sound reproducing device has to be a
during reproduction, for example for interruptions or for a repe
naturally involves the danger of dislocation.

As an alternative to this solution, the two lower adjustment
widened and elongated in the direction away from the other ad
toward the outside or in the direction of the same toward the in
respective aligning foot to be inserted into them with large clea
wide region created in this manner, so that said aligning foot i
narrow region and the other aligning foot/feet is/are in a positio
adjustment perforation/perforations in this position. A plane gu
this manner which passes the sound reproducing device place
round perforations of the sound carrier.

The walls of the enlarged adjustment perforations should h
configuration on the side facing toward or away from the other
as the aligning feet to be inserted. As a result, when centered,
in the exactly defined positions in which they would also be if
as is known, circular. The adjustment perforations not made la
common diameter, namely one of about 5-6 mm. The enlarge
nearly their halves the same configuration from where they wi
reach a larger transverse dimension of about 9-10 mm. The a
open toward the underside of the sound carrier. Actually, two
which one is circular and the other funnel-shaped are sufficien
Nevertheless, mostly four perforations are provided, as has be
which two are widened to funnel shape.

surround it a small distance of about 1.0 to 1.5 mm , especially 1.2 mm , thus forming a flat shoulder which contributesto defining the perforation	1-1	99.28%	einen geringen Abstand von etwa 1,0 bis 1,5 mm , insbesondere 1,2 mm , und bildet so einen flachen Absatz , was zur Präzisierung der Durchbrechung beiträgt
the annular bead should have a height of 0.2 to 0.3 mm , especially of 0.2 mm , and a width ofabout 1 -1.5 mm , especially 1.2 mm . it is advisable to provide the two perforations in the lower edge of the sound carrier with such an annular bead	1-2	48.85%	Der Wulstring sollte eine Höhe von 0,2 bis 0,3 mm , insbesondere von 0,2 mm , und eine Breite von ca . 1 - 1,5 mm , insbesondere 1,2 mm haben . Zweckmässigerweise werden zwei Durchbrüche am unteren Rand des Tonträgers mit diesem Wulstring ausgestattet
such anannular bead does not only make the perforations better visible , they also render the alignment of the feet of the sound reproducingdevice easier because they practically fall into the larger annular bead and slip into the precise adjustment perforationsautomatically	1-1	90.33%	Durch diesen Wulstring sind die Durchbrüche nicht nur besser sichtbar , sie erleichtern auch das Justieren der Füsse des Tonwiedergabegerätes , da die Justierfüsse in dem grösseren Wulstring praktisch hineinfallen und zwangsweise in die präzisen Justierdurchbrechungen schlüpfen
such a perforation surrounded by an annular bead can actually be compared to a funnel	1-1	21.93%	Man kann diese von einem Wulstring umgebenen Durchbrüche praktisch mit einem Trichter vergleichen .
of course , all the perforations in a sound carrier can be provided with beads	1-1	30.73%	Selbstverständlich können auch alle Durchbrechungen eines Tonträgers mit Wulsten versehen werden
they also prevent an unintentional dislocationof the sound reproducing device during possibly desired interruptions of a reproduction while studying educational information.This is important as the sound reproducing device has to be actuated frequently also during reproduction , for example forinterruptions or for a repeated reproduction which naturally involves the danger of dislocation	1-2	50.66%	Diese Wülste verhindern auch das unbeabsichtigte Verrücken des Tonwiedergabegerätes bei evtl . gewollten Unterbrechungen der Wiedergabe beim Studium von Lerntexten Das ist wichtig , weil das Tonwiedergabegerät auch während der Wiedergabe häufig , beispielsweise für Unterbrechungen oder Wiedergabewiederholungen zu betätigen ist und dabei an sich die Gefahr des Verrückens besteht .
as an alternative to this solution , the two lower adjustment perforations may be widened and elongated in the direction awayfrom the other adjustment perforations toward the outside or in the direction of the same toward the inside and surround therespective aligning foot to be inserted into them with large clearance by means of the wide region created in this manner , so that said aligning foot is movable to abut the narrow region and the other aligning foot / feet is / are in a position to engagethe other adjustment perforation / perforations in this position	1-1	100%	Alternativ zu dieser Lösung kann auch ein Teil der Justierdurchbrechungen , z . B. zwei untere Justierdurchbrechungen , in Richtung von den anderen Justierdurchbrechungen nach aussen weg oder zu diesen nach innen hin verbreitert und verlängert ausgebildet sein und den in sie einzuführenden Justierfuss mit dem so geschaffenen breiten Bereich mit grossem Spiel so umschliessen , dass dieser Justierfuss zur Anlage an den engen Bereich verschieblich ist und die anderen Justierfüsse in dieser Lage in die anderen Justierdurchbrechungen , die die Justierfüsse im engsten Bereich mit geringem Spiel umschliessen , einzugreifen vermögen
a plane guiding funnel is formed in this manner which passesthe sound reproducing device placed in position into the round perforations of the sound carrier	1-1	99.86%	So ist ein ebener Führungstrichter gebildet , der das aufgesetzte Tonwiedergabegerät in die runden Durchbrechungen des Tonträgers leitet
	0-1		Zusammen mit den dann zum Eingriff kommenden anderen Justierfüssen ist das Tonwiedergabegerät dann lagestabil gehalten .
	0-1		Ein älterer bekannter Vorschlag (US Patentschrift 4,298,967) ist von der Praxis als ungeeignet verworfen worden , der im Mittelpunkt der Tonrille eines folienartigen , dünnen Tonträgers eine schuhartige zur Mitte und nach unten sich verengende Vertiefung vorsah , in die ein zentraler Justierstift des Tonwiedergabegerätes bis zum Grund in die Justierstellung geführt wird , in diese aber wegen Fehlen der Arretierungsmöglichkeit nicht lagestabil festgehalten ist
	0-1		Ausserdem verdeckt das Tonwiedergabegerät jegliche Sicht zur schuhartigen Vertiefung .
the walls of the enlarged adjustment perforations should have the same configuration on the side facing toward or away fromthe other adjustment perforations as the aligning feet to be inserted	1-1	57.12%	Die Wände der vergrösserten Justierdurchbrechungen sollten auf der den anderen Justierdurchbrechungen zugewandten oder abgewandten Seite die gleiche Gestalt haben wie die einzusetzenden Justierfüsse
as a result , when centered , the feet are in each casein the exactly defined			Dadurch befinden sich die Füsse lagestabil in der zentrierten Lage jeweils an den genau

[0009] Zur Lösung dieses Problems sieht die Erfindung vor, dass bei dem eingangs
genannten Tonträger wenigstens eine der Justierdurchbrechungen auf der Oberseite
des Tonträgers von einem nach oben aus der Oberseite vorstehenden Wulstring, der z.
ist. Dieser Wulstring kann sich unmittelbar
lässt einen geringen Abstand von etwa 1,0
et so einen flachen Absatz, was zur
er Wulstring sollte eine Höhe von 0,2 bis 0,3
reite von ca. 1 - 1,5 mm, insbesondere 1,2

urchbrüche am unteren Rand des
et. Durch diesen Wulstring sind die
erleichtern auch das Justieren der Füße des
n dem größeren Wulstring praktisch
n Justierdurchbrechungen schlüpfen. Man
Durchbrüche praktisch mit einem Trichter

Durchbrechungen eines Tonträgers mit
rhindern auch das unbeabsichtigte
evtl. gewollten Unterbrechungen der
Das ist wichtig, weil das
edergabe häufig, beispielsweise für
lungen zu betätigen ist und dabei an sich

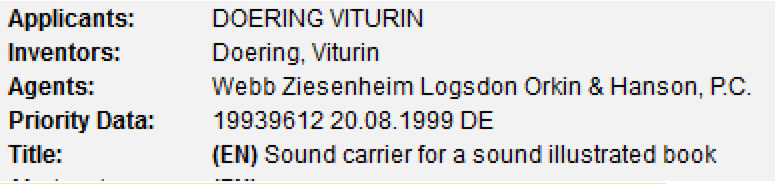
ch ein Teil der Justierdurchbrechungen, z.
Richtung von den anderen
der zu diesen nach innen hin verbreitert und
inzuführenden Justierfuß mit dem so
piel so umschließen, dass dieser Justierfuß
lich ist und die anderen Justierfüße in
hungen, die die Justierfüße im engsten
nzugreifen vermögen. So ist ein ebener
te Tonwiedergabegerät in die runden
sammen mit den dann zum Eingriff
onwiedergabegerät dann lagestabil

Patentschrift 4,298,967) ist von der Praxis
ittelpunkt der Tonrille eines folienartigen,
tte und nach unten sich verengende
rstift des Tonwiedergabegerätes bis zum
diese aber wegen Fehlen der
gehalten ist. Außerdem verdeckt das
hartigen Vertiefung.

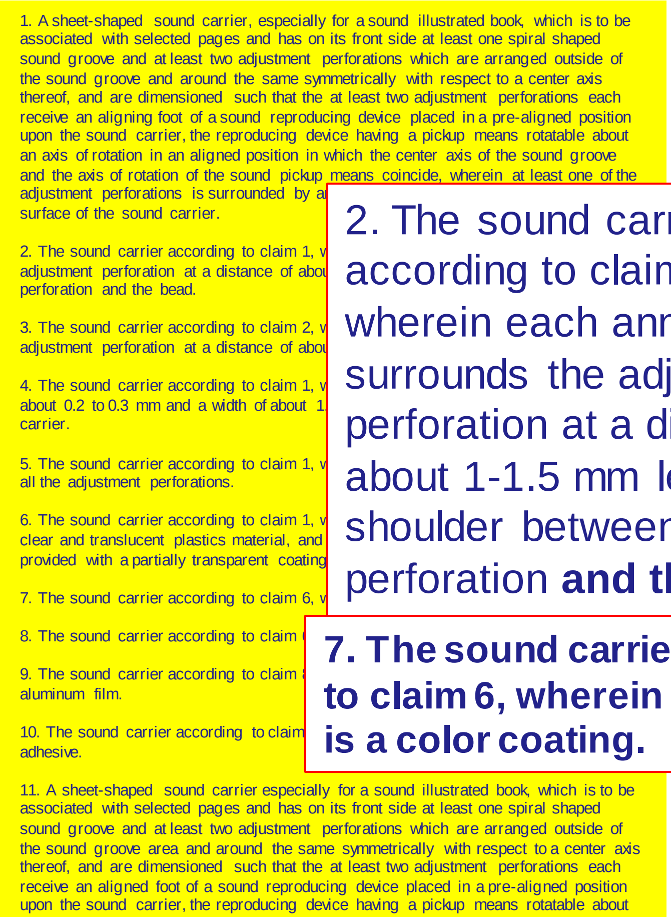
durchbrechungen sollten auf der den
ten oder abgewandten Seite die gleiche
erfüße. Dadurch befinden sich die Füße
den genau definierten Positionen, an denen
rchbrechungen, wie bekannt, kreisrund
rechungen sollten den üblichen

Claims alignment

Challenges with getting parallel sentences from claims



Applicants:	DOERING VITURIN
Inventors:	Doering, Viturin
Agents:	Webb Ziesenheim Logsdon Orkin & Hanson, P.C.
Priority Data:	19939612 20.08.1999 DE
Title:	(EN) Sound carrier for a sound illustrated book



1. A sheet-shaped sound carrier, especially for a sound illustrated book, which is to be associated with selected pages and has on its front side at least one spiral shaped sound groove and at least two adjustment perforations which are arranged outside of the sound groove and around the same symmetrically with respect to a center axis thereof, and are dimensioned such that the at least two adjustment perforations each receive an aligning foot of a sound reproducing device placed in a pre-aligned position upon the sound carrier, the reproducing device having a pickup means rotatable about an axis of rotation in an aligned position in which the center axis of the sound groove and the axis of rotation of the sound pickup means coincide, wherein at least one of the adjustment perforations is surrounded by a

surface of the sound carrier.

2. The sound carrier according to claim 1, w
adjustment perforation at a distance of abou
perforation and the bead.

3. The sound carrier according to claim 2, w
adjustment perforation at a distance of abou

4. The sound carrier according to claim 1, w
about 0.2 to 0.3 mm and a width of about 1.
carrier.

5. The sound carrier according to claim 1, w
all the adjustment perforations.

6. The sound carrier according to claim 1, w
clear and translucent plastics material, and
provided with a partially transparent coating

7. The sound carrier according to claim 6, w

8. The sound carrier according to claim 6

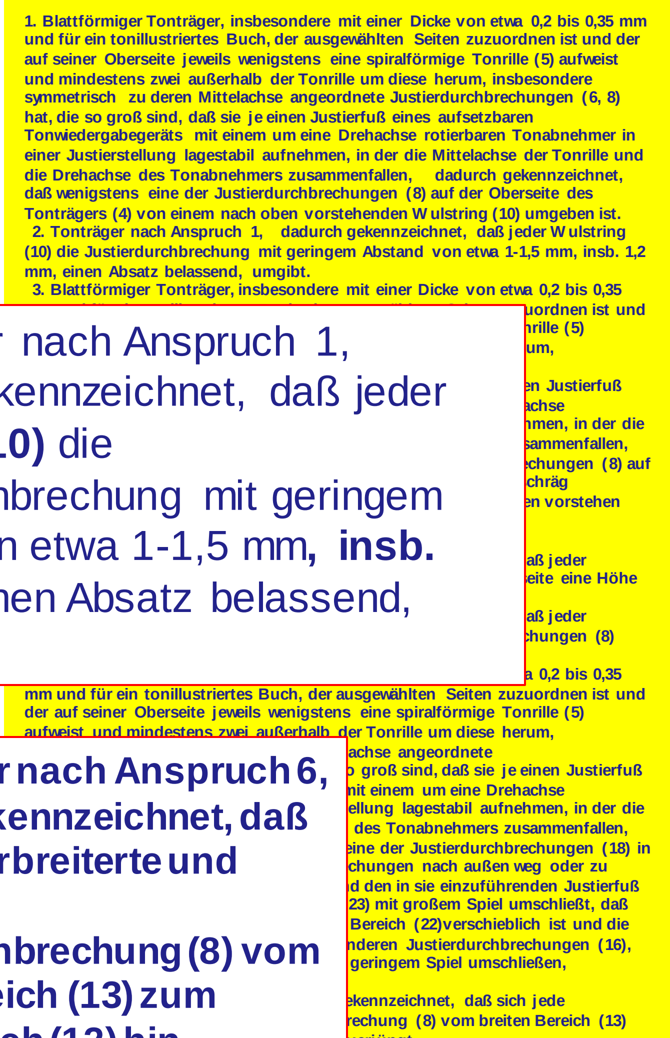
9. The sound carrier according to claim 8
aluminum film.

10. The sound carrier according to claim
adhesive.

11. A sheet-shaped sound carrier especially for a sound illustrated book, which is to be associated with selected pages and has on its front side at least one spiral shaped sound groove and at least two adjustment perforations which are arranged outside of the sound groove area and around the same symmetrically with respect to a center axis thereof, and are dimensioned such that the at least two adjustment perforations each receive an aligned foot of a sound reproducing device placed in a pre-aligned position upon the sound carrier, the reproducing device having a pickup means rotatable about

21 claims

16 claims



1. Blattförmiger Tonträger, insbesondere mit einer Dicke von etwa 0,2 bis 0,35 mm und für ein tonillustriertes Buch, der ausgewählten Seiten zuzuordnen ist und der auf seiner Oberseite jeweils wenigstens eine spiralförmige Tonrille (5) aufweist und mindestens zwei außerhalb der Tonrille um diese herum, insbesondere symmetrisch zu deren Mittelachse angeordnete Justierdurchbrechungen (6, 8) hat, die so groß sind, daß sie je einen Justierfuß eines aufsetzbaren Tonwiedergabegeräts mit einem um eine Drehachse rotierbaren Tonabnehmer in einer Justierstellung lagestabil aufnehmen, in der die Mittelachse der Tonrille und die Drehachse des Tonabnehmers zusammenfallen, dadurch gekennzeichnet, daß wenigstens eine der Justierdurchbrechungen (8) auf der Oberseite des Tonträgers (4) von einem nach oben vorstehenden W ulstring (10) umgeben ist.

2. Tonträger nach Anspruch 1, dadurch gekennzeichnet, daß jeder W ulstring (10) die Justierdurchbrechung mit geringem Abstand von etwa 1-1,5 mm, insb. 1,2 mm, einen Absatz belassend, umgibt.

3. Blattförmiger Tonträger, insbesondere mit einer Dicke von etwa 0,2 bis 0,35
uordnen ist und
nrille (5)
um,

en Justierfuß
achse
nmen, in der die
sammenfallen,
echungen (8) auf
chräg
en vorstehen

aß jeder
eite eine Höhe

aß jeder
chungen (8)

a 0,2 bis 0,35
mm und für ein tonillustriertes Buch, der ausgewählten Seiten zuzuordnen ist und der auf seiner Oberseite jeweils wenigstens eine spiralförmige Tonrille (5) aufweist und mindestens zwei außerhalb der Tonrille um diese herum,
achse angeordnete
o groß sind, daß sie je einen Justierfuß
mit einem um eine Drehachse
ellung lagestabil aufnehmen, in der die
des Tonabnehmers zusammenfallen,
eine der Justierdurchbrechungen (18) in
chungen nach außen weg oder zu
d den in sie einzuführenden Justierfuß
23) mit großem Spiel umschließt, daß
Bereich (22)verschieblich ist und die
nderen Justierdurchbrechungen (16),
geringem Spiel umschließen,

ekennzeichnet, daß sich jede
rechung (8) vom breiten Bereich (13)

2. The sound carrier according to claim 1, wherein each annular bead surrounds the adjustment perforation at a distance of about 1-1.5 mm leaving a shoulder between the perforation **and the bead**.

2. Tonträger nach Anspruch 1, dadurch gekennzeichnet, daß jeder Wulstring **(10)** die Justierdurchbrechung mit geringem Abstand von etwa 1-1,5 mm**, insb. 1,2 mm**, einen Absatz belassend, umgibt.

7. The sound carrier according to claim 6, wherein the coating is a color coating.

7. Tonträger nach Anspruch 6, dadurch gekennzeichnet, daß sich jede verbreiterte und verlängerte Justierdurchbrechung (8) vom breiten Bereich (13) zum engen Bereich (12) hin

Big models: bitexts

Matching Chinese description/claims with US

~ 64 Million segments (en zh)

English size: 2’000 Million words, 10Gb

(more careful alignment/cleaning to be done in the future)

Big models: language models

English texts only ~ 1Tb (US+EPO+PCT descriptions/claims/titles/abstracts)

…all English Wikipedia is 44Gb

Currently we stick to 10Gb for the language model

Tapta and big models

Parallelization:

- Use mgiza
- For big models: split corpus in 4 parts, launch mgiza on each quarter
- For big models: stop at HMM iteration

Heavy compression

- Pruning
- Binarization (compact phrase table, kenlm)
- (without much loss in quality)

Training and scalability
Size reduction zh-en

	Phrase table 0-0		Phrase table 0,1-0		Reordering model		Language model	
	M rows	Gb	M rows	Gb	M rows	Gb	M ngrams	Gb
Basic	806	100	974	130	806	89	584	23
Pruned	551	69	623	83	551	61	388	16
Binarized		6.4		7.4		4.2		4.6

342Gb

↓

22.6Gb (6.6%)

Language specificities

Tokenizer: Based on Lucene framework

- zh: adapted “SmartCn”
- ja: “kuromoji”
- de: decompounder (Junczys & Pouliquen, Eamt2014)
- ar: prefix splitting, removes shot vowels
- ko: decompounder
- Normalizes greek letters
- Groups references to figures (eg. “(1)” not “_(_1_)_”)

Reordering

- de: pre-reordering (Junczys & Pouliquen, Eamt2014)
- ja: Simple naïve pre-reordering (more to be done)

Tapta: our tool installed in different places

WIPO PATENTSCOPE

WIPO | PCT
The International
Patent System

WIPO MADRID
The International Trademark System

International
Telecommunication
Union

Tapta4UN
United Nations
New York

And some projects going on to install Tapta prototype in other International institutions…

Our tool in different situations

- Adapted our code so that it can easily install and run
- Under version control, regression tests, Installation/administration documentation (100 pages)
- With installation instructions: ½ day to configure a new Linux server
- Runs on Linux:
  - Hardware: Amazon cloud, virtual machine, desktop PC, server
  - OS: Ubuntu/Suse/Centos/RedHat

Training and scalability: UN data

- All United Nation texts ~ 212 Million words, 10 M segments

	Phrase table		Reordering model		Language model	
	M rows	Gb	M rows	Gb	M rows	Gb
Basic	82	9.70	82	8.70	49	1.70
Pruned	19	2.20	19	1.90	31	1.00
Binarized		0.27		0.15		0.70

UN data

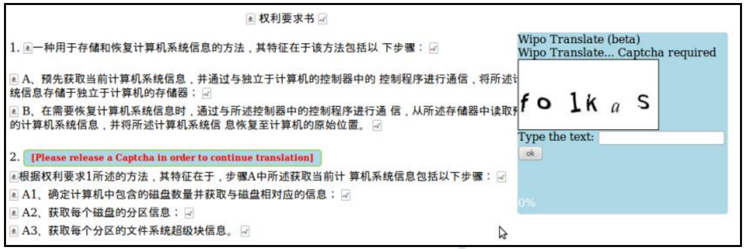
20Gb

↓

1.12Gb (6%)

Our tool in production

- Install/publish/update/train/evaluate (robust) scripts
- Monitor tool
- Dashboard interface
- Anti robot policy (captcha)
- Statistics
- …



权利要求书

1. 一种用于存储和恢复计算机系统信息的方法，其特征在于该方法包括以 下步骤：

A、预先获取当前计算机系统信息，并通过与独立于计算机的控制器中的 控制程序进行通信，将所述

统信息存储于独立于计算机的存储器；

B、在需要恢复计算机系统信息时，通过与所述控制器中的控制程序进行通 信，从所述存储器中读取

的计算机系统信息，并将所述计算机系统信 息恢复至计算机的原始位置。

2. [Please release a Captcha in order to continue translation]

根据权利要求1所述的方法，其特征在于，步骤A中所述获取当前计 算机系统信息包括以下步骤：

A1、确定计算机中包含的磁盘数量并获取与磁盘相对应的信息；

A2、获取每个磁盘的分区信息；

A3、获取每个分区的文件系统超级块信息。

Wipo Translate (beta)
Wipo Translate... Captcha required

Type the text:

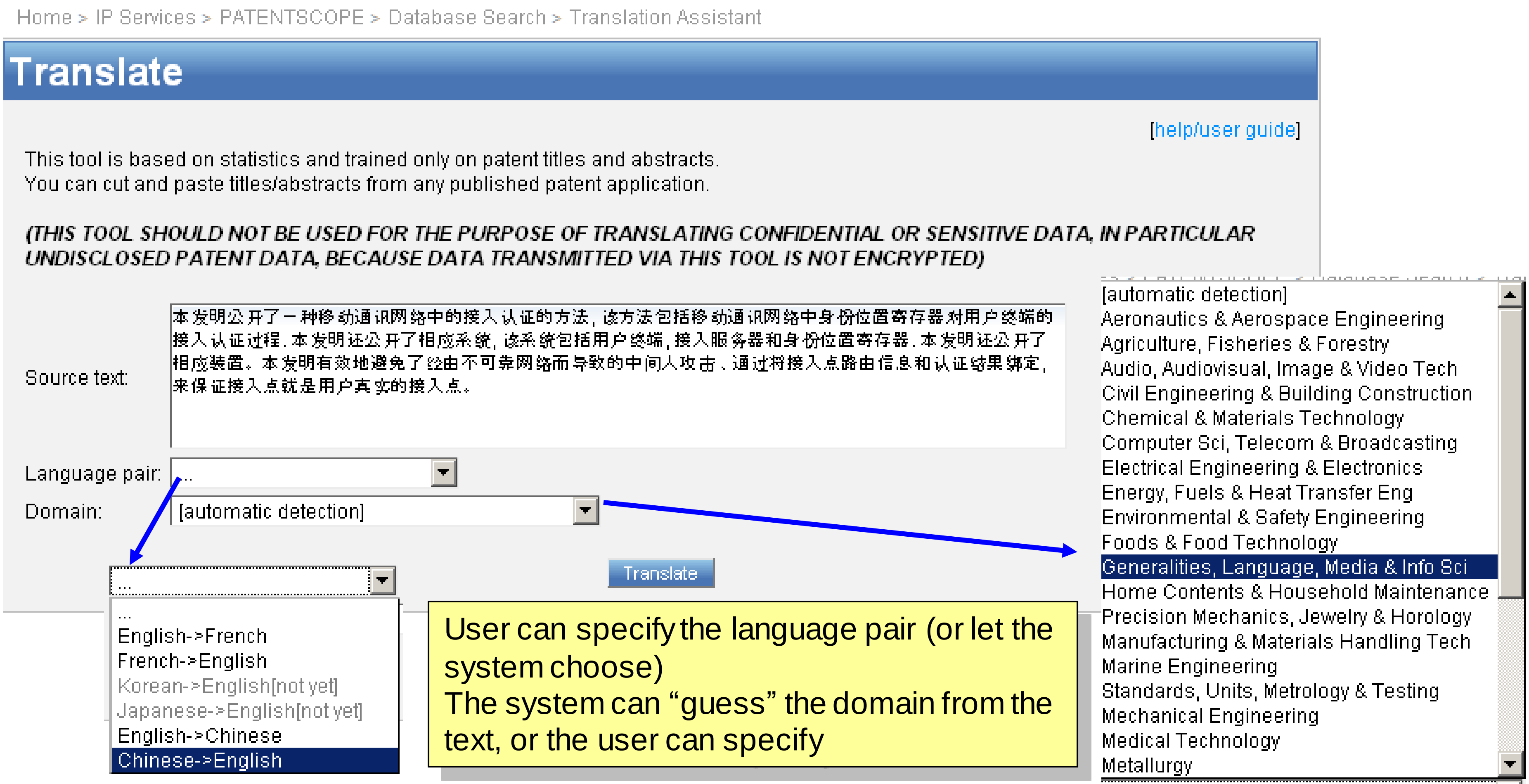
Hardware & OS

- Virtual, 4 Gb Ram, Suse SLES11, 4 cores, 250Gb
- Virtual, 16 Gb Ram, RedHat ent R6.2, 16 cores, 200Gb
- PC, 8Gb ram, Ubuntu 12.4, 8 cores, 350Gb
- Server, 16Gb, Centos R6.4, 16 cores, 400Gb
- Server, 11Gb Ram, RedHat ent. R6.5, 8 cores, 100Gb disk
- …
- Amazon cloud, 64 Gb Ram, Suse ent. 11, 8 cores, 400Gb
- …
- Server, 500Gb ram, RedHat ent. R6.5, 48 cores, 4T

Various user interfaces

- Java Swing
- Web interface
  - Gist translation
  - Interactive translation
- Hotkey
- Plugin (SDL studio, Worldserver, eLuna etc.)
- Tapta widget

Web interface



Home > IP Services > PATENTSCOPE > Database Search > Translation Assistant

Translate

[help/user guide]

This tool is based on statistics and trained only on patent titles and abstracts.
You can cut and paste titles/abstracts from any published patent application.

(THIS TOOL SHOULD NOT BE USED FOR THE PURPOSE OF TRANSLATING CONFIDENTIAL OR SENSITIVE DATA, IN PARTICULAR UNDISCLOSED PATENT DATA, BECAUSE DATA TRANSMITTED VIA THIS TOOL IS NOT ENCRYPTED)

Source text: 本发明公开了一种移动通讯网络中的接入认证的方法, 该方法包括移动通讯网络中身份位置寄存器对用户终端的接入认证过程. 本发明还公开了相应系统, 该系统包括用户终端, 接入服务器和身份位置寄存器. 本发明还公开了相应装置。本发明有效地避免了经由不可靠网络而导致的中间人攻击、通过将接入点路由信息和认证结果绑定, 来保证接入点就是用户真实的接入点。

Language pair: ...

Domain: [automatic detection]

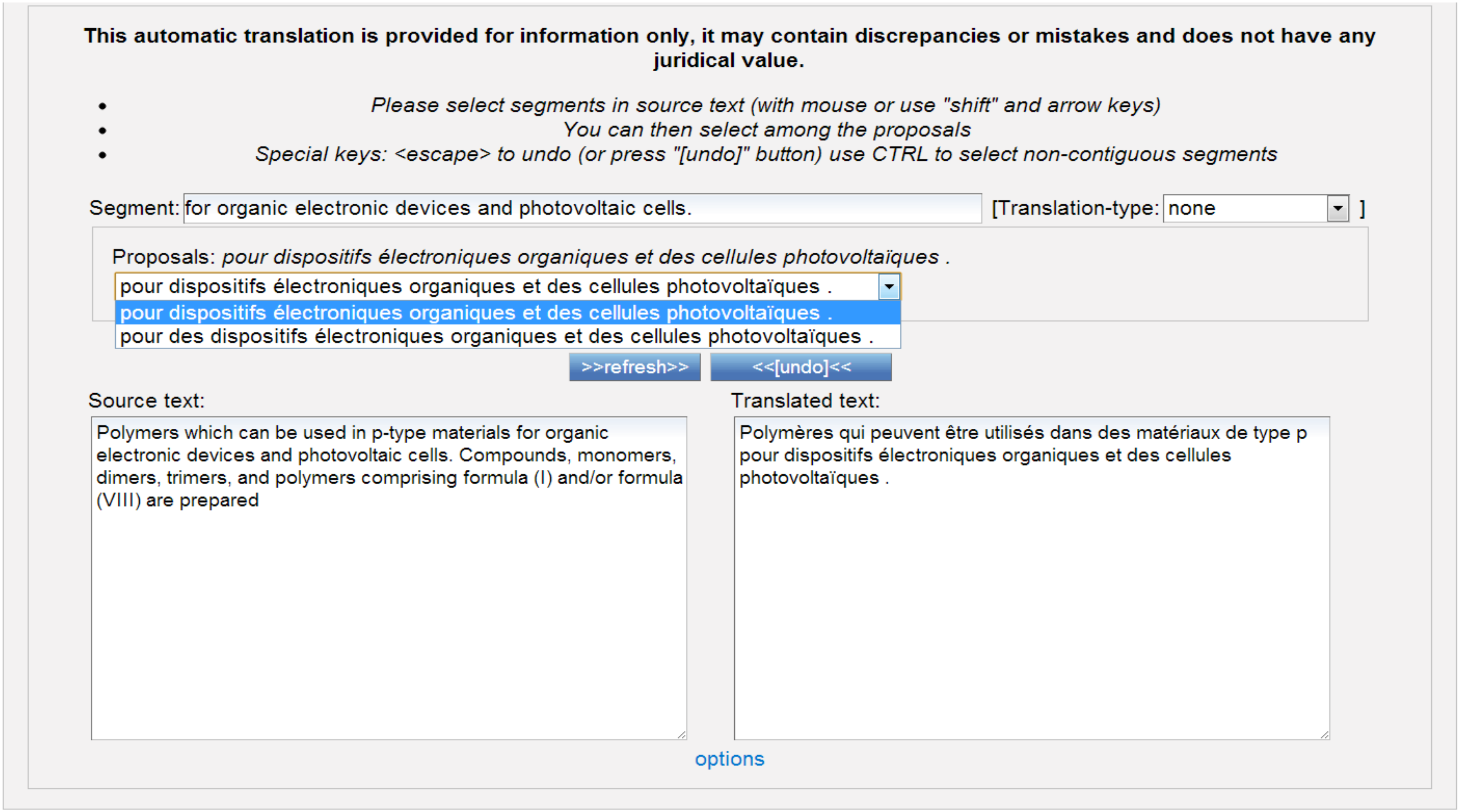
Translate

...
English->French
French->English
Korean->English[not yet]
Japanese->English[not yet]
English->Chinese
Chinese->English

[automatic detection]
Aeronautics & Aerospace Engineering
Agriculture, Fisheries & Forestry
Audio, Audiovisual, Image & Video Tech
Civil Engineering & Building Construction
Chemical & Materials Technology
Computer Sci, Telecom & Broadcasting
Electrical Engineering & Electronics
Energy, Fuels & Heat Transfer Eng
Environmental & Safety Engineering
Foods & Food Technology
Generalities, Language, Media & Info Sci
Home Contents & Household Maintenance
Precision Mechanics, Jewelry & Horology
Manufacturing & Materials Handling Tech
Marine Engineering
Standards, Units, Metrology & Testing
Mechanical Engineering
Medical Technology
Metallurgy

User can specify the language pair (or let the system choose)
The system can “guess” the domain from the text, or the user can specify

TAPTA web interactive



This automatic translation is provided for information only, it may contain discrepancies or mistakes and does not have any juridical value.

- *Please select segments in source text (with mouse or use "shift" and arrow keys)*
- *You can then select among the proposals*
- *Special keys: <escape> to undo (or press "[undo]" button) use CTRL to select non-contiguous segments*

Segment: for organic electronic devices and photovoltaic cells. [Translation-type: none]

Proposals: *pour dispositifs électroniques organiques et des cellules photovoltaïques .*

pour dispositifs électroniques organiques et des cellules photovoltaïques .

pour dispositifs électroniques organiques et des cellules photovoltaïques .

pour des dispositifs électroniques organiques et des cellules photovoltaïques .

Source text:

Polymers which can be used in p-type materials for organic electronic devices and photovoltaic cells. Compounds, monomers, dimers, trimers, and polymers comprising formula (I) and/or formula (VIII) are prepared

Translated text:

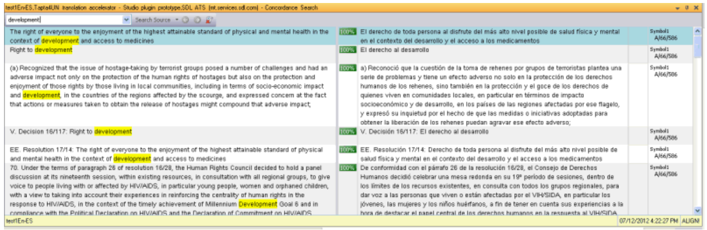
Polymères qui peuvent être utilisés dans des matériaux de type p pour dispositifs électroniques organiques et des cellules photovoltaïques .

Tapta hotkey

- Access the translation server using the "F3" key: Select text, press F3, translation goes to clipboard
- Work only on PC (opensource AutoHotKey), but is a solution to integrate MT in any application

Plugins

SDL/Studio



test1En-ES.Tapta4UN translation accelerator - Studio plugin prototype.SDL ATS (mt.services.sdl.com) - Concordance Search

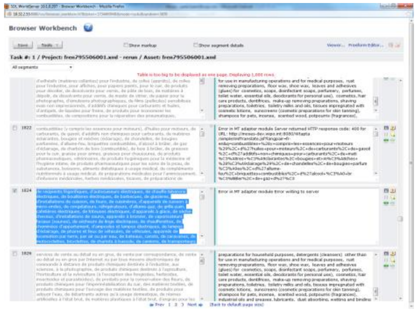
development

The right of everyone to the enjoyment of the highest attainable standard of physical and mental health in the context of development and access to medicines	100%	El derecho de toda persona al disfrute del más alto nivel posible de salud física y mental en el contexto del desarrollo y el acceso a los medicamentos	Symbol1 A/66/586
Right to development	100%	El derecho al desarrollo	Symbol1 A/66/586
(a) Recognized that the issue of hostage-taking by terrorist groups posed a number of challenges and had an adverse impact not only on the protection of the human rights of hostages but also on the protection and enjoyment of those rights by those living in local communities, including in terms of socio-economic impact and development, in the countries of the regions affected by the scourge, and expressed concern at the fact that actions or measures taken to obtain the release of hostages might compound that adverse impact;	100%	a) Reconoció que la cuestión de la toma de rehenes por grupos de terroristas plantea una serie de problemas y tiene un efecto adverso no solo en la protección de los derechos humanos de los rehenes, sino también en la protección y el goce de los derechos de quienes viven en comunidades locales, en particular en términos de impacto socioeconómico y de desarrollo, en los países de las regiones afectadas por ese flagelo, y expresó su inquietud por el hecho de que las medidas o iniciativas adoptadas para obtener la liberación de los rehenes puedan agravar ese efecto adverso;	Symbol1 A/66/586
V. Decision 16/117: Right to development	100%	V. Decisión 16/117: El derecho al desarrollo	Symbol1 A/66/586
EE. Resolution 17/14: The right of everyone to the enjoyment of the highest attainable standard of physical and mental health in the context of development and access to medicines	100%	EE. Resolución 17/14: Derecho de toda persona al disfrute del más alto nivel posible de salud física y mental en el contexto del desarrollo y el acceso a los medicamentos	Symbol1 A/66/586
70. Under the terms of paragraph 26 of resolution 16/28, the Human Rights Council decided to hold a panel discussion at its nineteenth session, within existing resources, in consultation with all regional groups, to give voice to people living with or affected by HIV/AIDS, in particular young people, women and orphaned children, with a view to taking into account their experiences in reinforcing the centrality of human rights in the response to HIV/AIDS, in the context of the timely achievement of Millennium Development Goal 6 and in compliance with the Political Declaration on HIV/AIDS and the Declaration of Commitment on HIV/AIDS	100%	De conformidad con el párrafo 26 de la resolución 16/28, el Consejo de Derechos Humanos decidió celebrar una mesa redonda en su 19º período de sesiones, dentro de los límites de los recursos existentes, en consulta con todos los grupos regionales, para dar voz a las personas que viven o están afectadas por el VIH/SIDA, en particular los jóvenes, las mujeres y los niños huérfanos, a fin de tener en cuenta sus experiencias a la hora de destacar el papel central de los derechos humanos en la respuesta al VIH/SIDA	Symbol1 A/66/586

test1En-ES 07/12/2012 4:22:27 PM ALIGN!

SDL Trados Studio 2014

Worldserver



Browser Workbench

Task #: 1 / Project: fren795506001.xml - rerun / Asset: fren795506001.xml

All segments

Table is too big to be displayed as one page. Displaying 1,000 rows.

1822 Error in MT adapter module Server returned HTTP response code: 400 for URL: http://mmes-dev.wipo.int:8080/Wtapta

1824 Error in MT adapter module error writing to server

1826

SDL WorldServer

Eluna

(United Nation internal CAT tool)

Multitrans?

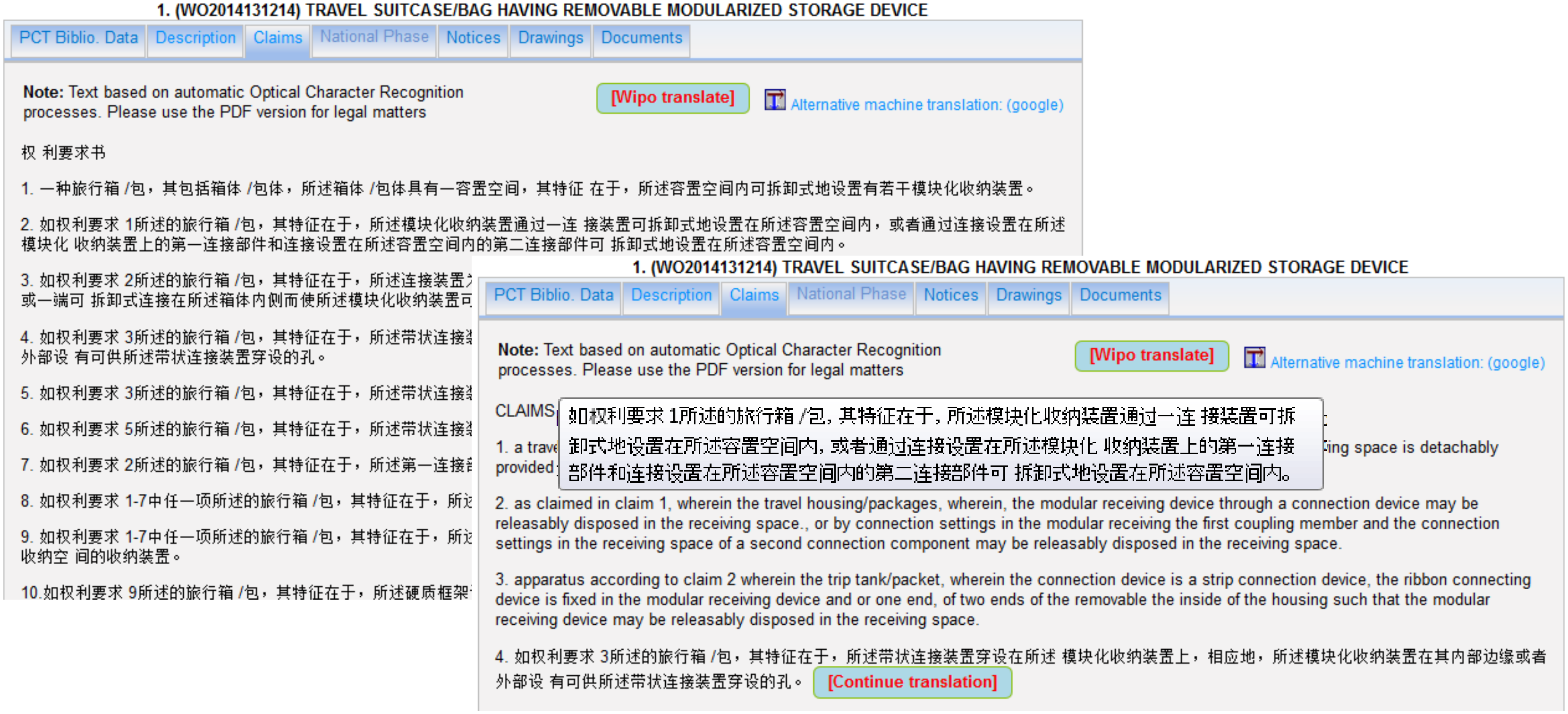
MultiTRANS Prism 5.6

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Tapta widget

Demo

A script to be inserted at the beginning of any HTML page, translates inline text on the fly



1. (WO2014131214) TRAVEL SUITCASE/BAG HAVING REMOVABLE MODULARIZED STORAGE DEVICE

Note: Text based on automatic Optical Character Recognition processes. Please use the PDF version for legal matters

权 利要求书

1. 一种旅行箱 /包，其包括箱体 /包体，所述箱体 /包体具有一容置空间，其特征 在于，所述容置空间内可拆卸式地设置有若干模块化收纳装置。

2. 如权利要求 1所述的旅行箱 /包，其特征在于，所述模块化收纳装置通过一连 接装置可拆卸式地设置在所述容置空间内，或者通过连接设置在所述模块化 收纳装置上的第一连接部件和连接设置在所述容置空间内的第二连接部件可 拆卸式地设置在所述容置空间内。

3. 如权利要求 2所述的旅行箱 /包，其特征在于，所述连接装置
或一端可 拆卸式连接在所述箱体内侧而使所述模块化收纳装置可

4. 如权利要求 3所述的旅行箱 /包，其特征在于，所述带状连接
外部设 有可供所述带状连接装置穿设的孔。

5. 如权利要求 3所述的旅行箱 /包，其特征在于，所述带状连接

6. 如权利要求 5所述的旅行箱 /包，其特征在于，所述带状连接

7. 如权利要求 2所述的旅行箱 /包，其特征在于，所述第一连接

8. 如权利要求 1-7中任一项所述的旅行箱 /包，其特征在于，所

9. 如权利要求 1-7中任一项所述的旅行箱 /包，其特征在于，所
收纳空 间的收纳装置。

10.如权利要求 9所述的旅行箱 /包，其特征在于，所述硬质框架

1. (WO2014131214) TRAVEL SUITCASE/BAG HAVING REMOVABLE MODULARIZED STORAGE DEVICE

Note: Text based on automatic Optical Character Recognition processes. Please use the PDF version for legal matters

CLAIMS

如权利要求 1所述的旅行箱 /包，其特征在于，所述模块化收纳装置通过一连 接装置可拆卸式地设置在所述容置空间内，或者通过连接设置在所述模块化 收纳装置上的第一连接部件和连接设置在所述容置空间内的第二连接部件可 拆卸式地设置在所述容置空间内。

1. a trave ing space is detachably provided

2. as claimed in claim 1, wherein the travel housing/packages, wherein, the modular receiving device through a connection device may be releasably disposed in the receiving space., or by connection settings in the modular receiving the first coupling member and the connection settings in the receiving space of a second connection component may be releasably disposed in the receiving space.

3. apparatus according to claim 2 wherein the trip tank/packet, wherein the connection device is a strip connection device, the ribbon connecting device is fixed in the modular receiving device and or one end, of two ends of the removable the inside of the housing such that the modular receiving device may be releasably disposed in the receiving space.

4. 如权利要求 3所述的旅行箱 /包，其特征在于，所述带状连接装置穿设在所述 模块化收纳装置上，相应地，所述模块化收纳装置在其内部边缘或者外部设 有可供所述带状连接装置穿设的孔。

Tapta: translation quality

- Competitive!

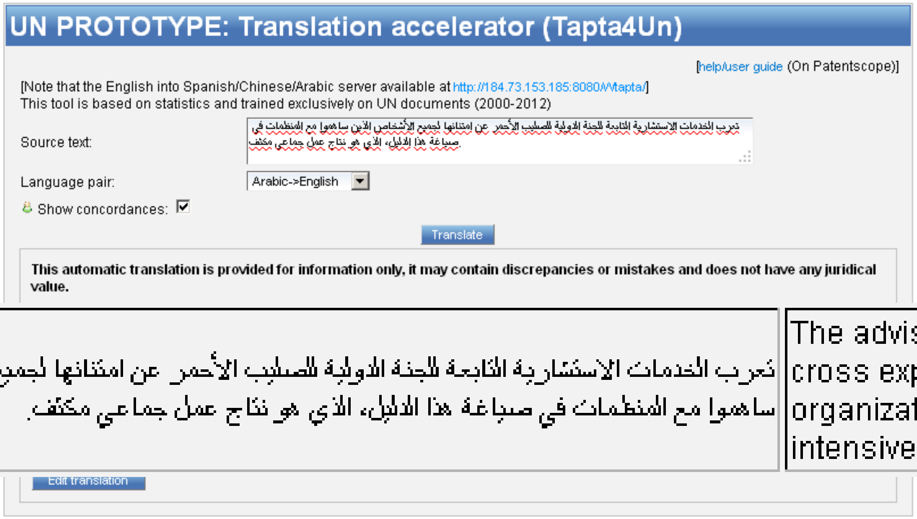
- Better than Google and Microsoft translate
  - Working with our data
  - Based on good open source “Moses”
- Small team…

  but working with others…

Also the United Nations

BLEU scores

Language pair	*Tapta*	*Google*	*Bing*
ar-en	55.25	n/a[1]	51.17
en-ar	44.10	33.74	28.94
en-es	61.81	53.39	46.86
en-fr	51.23	45.58	42.19
en-ru	50.85	39.67	38.96
en-zh	43.17	34.16	32.77
es-en	60.32	52.54	49.18
fr-en	53.36	46.46	43.39
ru-en	58.56	47.71	47.09
zh-en	42.31	36.55	30.60



UN PROTOTYPE: Translation accelerator (Tapta4Un)

[help/user guide (On Patentscope)]

[Note that the English into Spanish/Chinese/Arabic server available at http://184.73.153.185:8080/Wtapta/]
This tool is based on statistics and trained exclusively on UN documents (2000-2012)

Source text: تعرب الخدمات الاستشارية التابعة للجنة الدولية للصليب الأحمر عن امتنانها لجميع الأشخاص الذين ساهموا مع المنظمات في صياغة هذا الدليل، الذي هو نتاج عمل جماعي مكثف.

Language pair: Arabic->English

Show concordances: ☑

This automatic translation is provided for information only, it may contain discrepancies or mistakes and does not have any juridical value.

تعرب الخدمات الاستشارية التابعة للجنة الدولية للصليب الأحمر عن امتنانها لجميع الأشخاص الذين
ساهموا مع المنظمات في صياغة هذا الدليل، الذي هو نتاج عمل جماعي مكثف.

The advisory services of the international committee of the red cross expressed its gratitude to all those who have contributed with organizations in drafting the manual, which was the product of intensive teamwork.

Bruno Pouliquen, Cecilia Elizalde, Marcin Junczys-Dowmunt, Christophe Mazenc, José García-Verdugo, **Large-scale multiple language translation accelerator at the United Nations** , MT Summit, Nice, France, September 2013

User acceptance

How Tapta is perceived among translators:

- When seen as a “translation accelerator”: very useful
- When seen as “replacement for translator”: useless
- When proposed as a copy-paste tool: not used
- When integrated in translator’s environment: used

Frustration: User has little impact on the MT output
Blacklist that we apply on the phrase table
Collect post-edition segments:

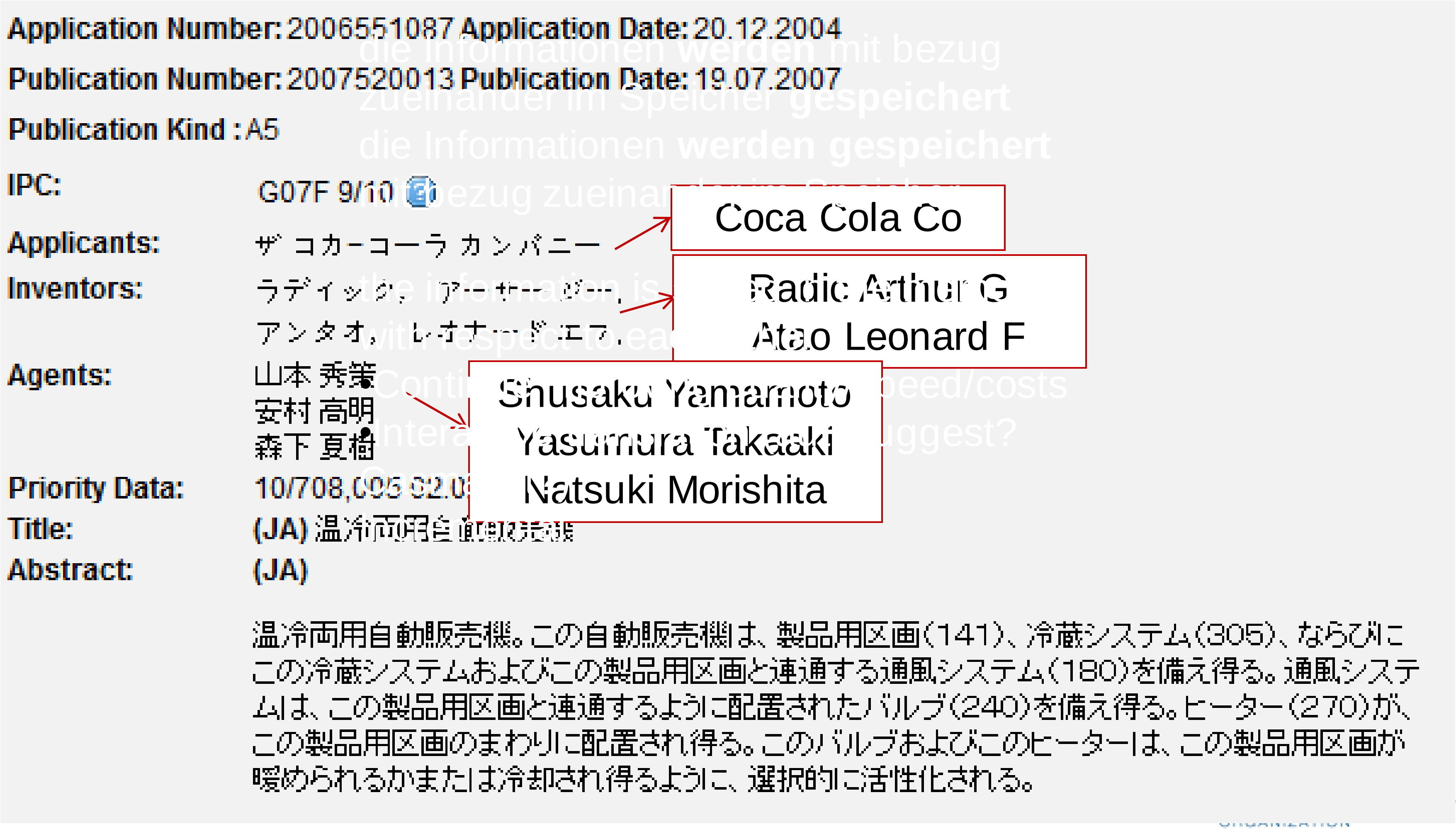
- quality estimation
- improving the MT

.

Conclusion/discussion

- MT contributes to information dissemination
- Moses easily supports huge models
- Tapta MT quality competitive
- Language dependent tools should be avoided in our context
- “User acceptance landscape is changing”
- Integration!

Future work on transliteration



Application Number: 2006551087 **Application Date:** 20.12.2004

Publication Number: 2007520013 **Publication Date:** 19.07.2007

Publication Kind : A5

IPC: G07F 9/10

Applicants: ザ コカ－コーラ カンパニー

Inventors: ラディック, アーサー ジー,
アンタオ, レオナード エフ.

Agents: 山本 秀策
安村 高明
森下 夏樹

Priority Data: 10/708,005 02.0

Title: (JA) 温冷両用自動販売機

Abstract: (JA)

温冷両用自動販売機。この自動販売機は、製品用区画(141)、冷蔵システム(305)、ならびにこの冷蔵システムおよびこの製品用区画と連通する通風システム(180)を備え得る。通風システムは、この製品用区画と連通するように配置されたバルブ(240)を備え得る。ヒーター(270)が、この製品用区画のまわりに配置され得る。このバルブおよびこのヒーターは、この製品用区画が暖められるかまたは冷却され得るように、選択的に活性化される。

Coca Cola Co

Radic Arthur G
Atao Leonard F

Shusaku Yamamoto
Yasumura Takaaki
Natsuki Morishita

Future work

- Continue improving quality/speed/costs
  - OSM
  - Word cluster LM
  - Additional usage of meta data
- Interactive translation (autosuggest?)
- Incremental training
- Translation through pivot language

Future work

User feedbacks

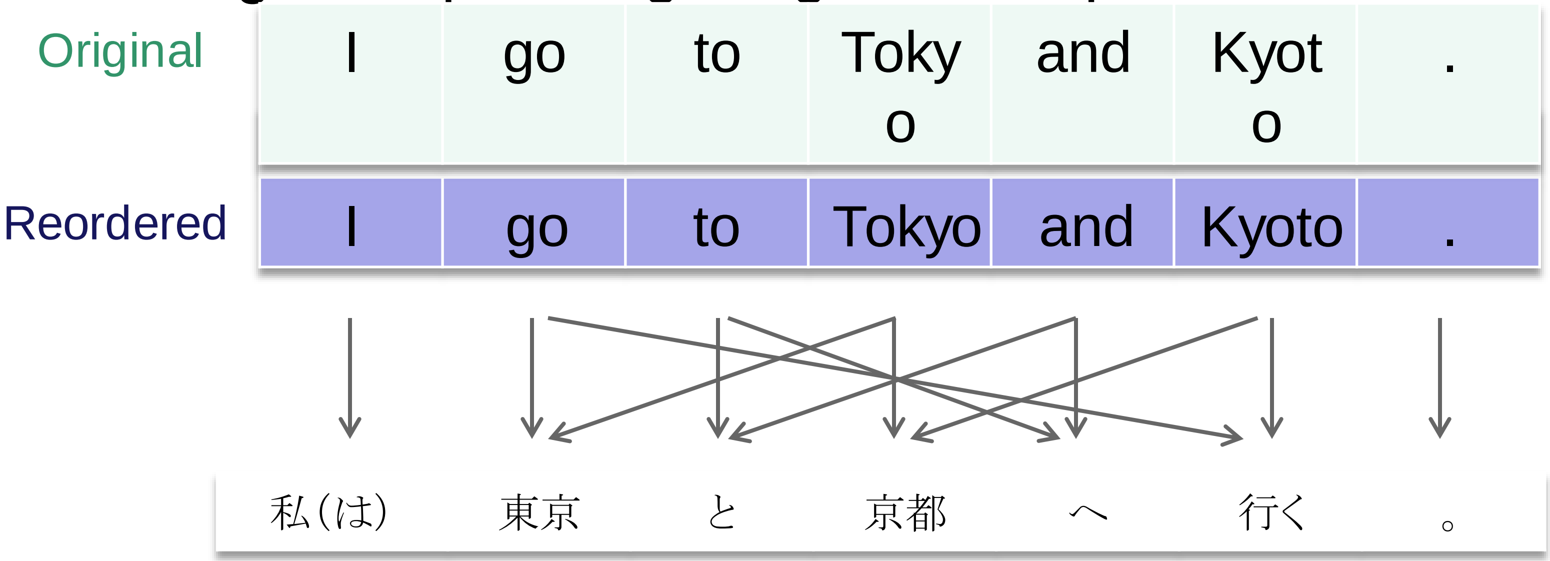
- Take into account new translations
- Blacklist of phrases
- Collect post-editions
- …

Pre-reordering

Dan Han
Native Chinese
PhD Tokyo Graduate University for Advanced Studies

Order differences complicate phrase extraction.

Reordering is helpful. E.g. English-to-Japanese



Original	I	go	to	Toky o	and	Kyot o	.
Reordered	I	go	to	Tokyo	and	Kyoto	.
	私(は)	東京	と	京都	へ	行く	。

Thank you for your attention

شكرا لكم على اهتمامكم

Merci pour votre attention!

感谢您的关注

Grazie per la vostra attenzione!

¡ Gracias por su atención !

Vielen Dank für Ihre Aufmerksamkeit!

Obrigado pela vossa atenção!

Dziękuję bardzo za Państwa uwagę!

Děkujeme za Vaši pozornost!

Ďakujem ti vel’mi pekne za tvoju pozornost’

Tänan tähelepanu eest!

Благодарим за Вашето внимание!

Tak for Jeres opmærksomhed!

आप अपना ध्यान के लिए धन्यवाद

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Bibliography

Try it! Google: wipo translate

(2014) Marcin Junczys-Dowmunt and Bruno Pouliquen: SMT of German Patents at WIPO: Decompounding and Verb Structure Pre-reordering. *(EAMT 2014)*, 16-18 June 2014, Dubrovnik

(2013) Bruno Pouliquen, Christophe Mazenc & Paul Halfpenny: Latest developments in machine translation at WIPO, *EPO- East meets West*, 18-19 April 2013, Vienna, Austria

(2011) Bruno Pouliquen & Christophe Mazenc: Automatic translation tools at WIPO. *Aslib, Translating and the Computer 33*, 17-18 Nov 2011, London

(2012) Marcin Junczys-Dowmunt : A Phrase Table without Phrases: Rank Encoding for Better Phrase Table Compression. *EAMT 2012*

(2011) Bruno Pouliquen & Christophe Mazenc: COPPA, CLIR and TAPTA: three tools to assist in overcoming the patent barrier at WIPO. *MT Summit XIII*.

(2011) Bruno Pouliquen, Christophe Mazenc & Aldo Iorio: Tapta: a user-driven translation system for patent documents based on domain-aware statistical machine translation. *[EAMT 2011]*

Tapta4UN:

(2013) B. Pouliquen, C. Elizalde, M. Junczys-Dowmunt, C. Mazenc & J. Garcia-Verdugo: Large-scale multiple language translation accelerator at the United Nations. [MT Summit 2013]

(2012) C. Elizalde, B. Pouliquen, C. Mazenc & J. García-Verdugo: TAPTA4UN: collaboration on machine translation between the World Intellectual Property Organization and the United Nations. [Aslib 2012] *Translating and the Computer 34*, 29-30 November 2012]

(2012) Bruno Pouliquen, Christophe Mazenc, Cecilia Elizalde, & Jose Garcia-Verdugo: Statistical machine translation prototype using UN parallel documents. *EAMT 2012*