



Advanced Technology Applications Center

strategy for developing AI projects, list of current AI projects

Bruno Pouliquen,
World Intellectual Property Organization (WIPO)

May/2018

Machine learning at WIPO

- WIPO had experience in:
 - Machine translation
 - Cross language Information retrieval
 - IPC classification
- New sector: Advanced Technology Applications Center (ATAC)
R&D center to explore various machine learning technologies in IP context
 - 2 permanent staff
 - 3 fellows : Machine Translation / Speech / Image classification
 - External provider support
 - (and collaboration with other Organizations...)

Machine learning projects

- Machine translation
- Speech to text
 - Speech to translated text
- Image classification
- Others:
 - Text-based classification (IPC, CPC, etc.)
 - Proper names (categorization, transliteration, cleaning...)
 - Chatbots
 - Etc.

Our strategy, what makes successful projects

- Technological survey, academic network
- Open source technologies (so that we can share our development with IPOs)
- Use “recent/cleaned/big” data
 - Usually thousands of “input/output”
 - Use machine learning to “guess” the best “output” knowing new “input”
- Careful choice of potential successful projects
- Quick prototyping / “fail fast to innovate faster”
- More exploring/cleaning/adaptation needed? → Discuss about budget and workforce



AI and Machine Translation at WIPO

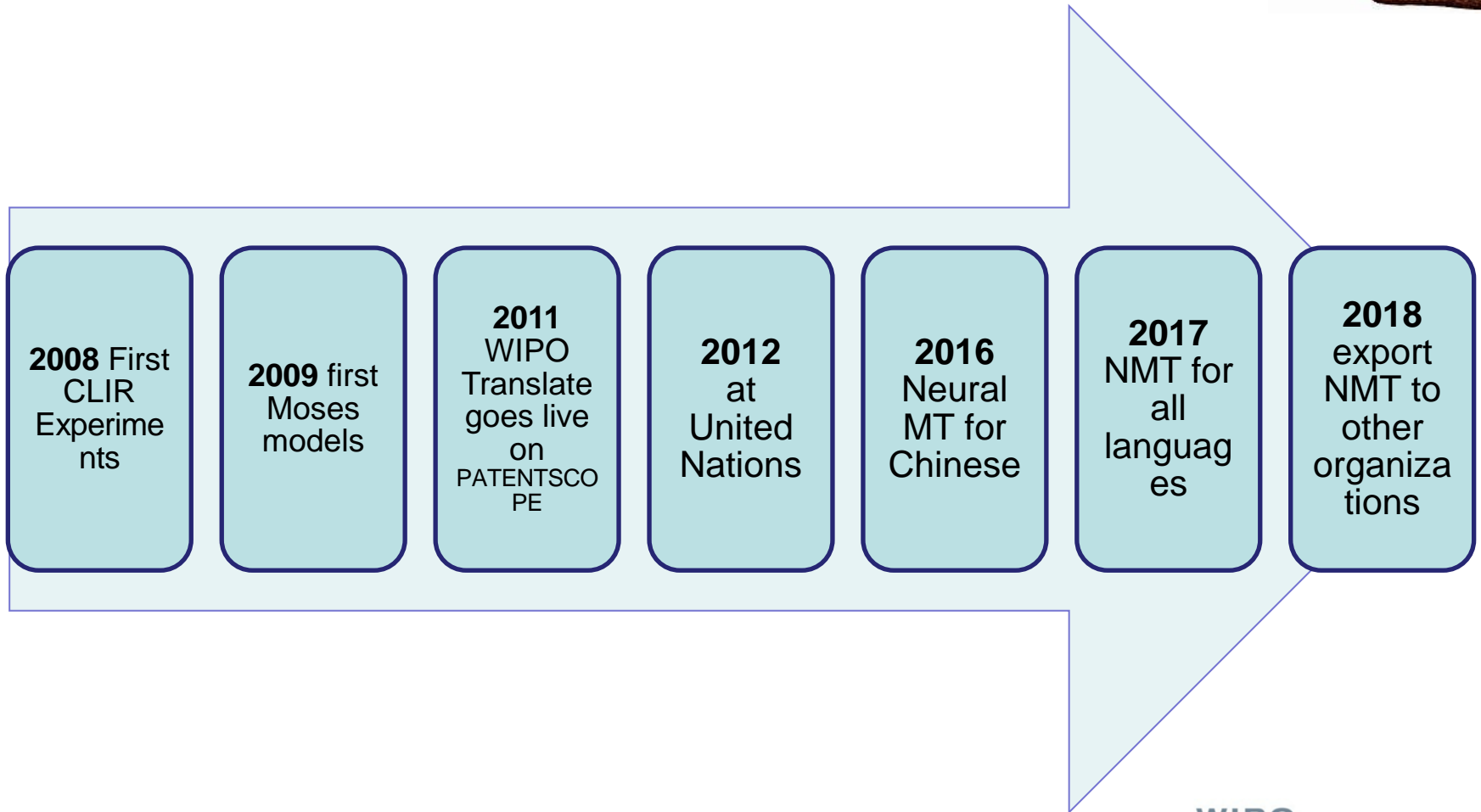
Bruno Pouliquen,
World Intellectual Property Organization (WIPO)

May/2018

What is WIPO Translate?

- Data-driven machine translation
- In-house development based on open source
- Neural Machine Translation (NMT) in production on PATENTSCOPE since September 2016
- All 10 PCT official languages use NMT
- Fast and accurate (outperforms other tools)
- Freely available to the public
- WIPO can share the software with other organizations

WIPO Translate history



WIPO Translate's goals



- Scalability:
 - Fast translations (on the fly), speed: <math><1\text{ s}</math> per sentence
 - Can absorb huge quantity of training data
- Usability:
 - Free to use (open source + in-house development)
 - Fully integrated (in PATENTSCOPE), but also as a standalone tool
 - Confidentiality (<https>)
- Quality:
 - Provides “state-of-the-art” quality (for in-domain texts)
 - Competitive for all language pairs

How do we build our training data

- Use translated title and abstracts
 - From PCT, from IPOs
- For Patents: use fulltext
 - Use priorities to access “comparable” descriptions and claims (SIPO application having US priority or vice-versa)
 - Align patent applications belonging to the same “family”
 - Align sentences between the two applications
 - ... and filter them!

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

[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.

To meet this object, the invention provides that in the sound carrier at least one of the adjustment perforations on the surface of the sound carrier is surrounded by an annular bead which protrudes in upward direction and has been impressed or deep-drawn, for example. Said annular bead is located immediately at the perforation or surround it a small distance, especially 1.2 mm, thus forming a flat shoulder which contributes to defining the perforation.

It is advisable to provide the two perforations in the lower edge of the sound carrier with such an annular bead. Such an annular bead does not only make the perforations better visible, they also render the alignment of the feet of the sound reproducing device easier because they practically fall into the larger annular bead and slip into the precise adjustment perforations automatically.

Of course, all the perforations in a sound carrier can be provided with beads. They also prevent an unintentional dislocation of 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 for interruptions or for a repeated reproduction which naturally involves the danger of dislocation.

As an alternative to this solution, the two lower adjustment perforations may be widened and elongated in the direction away from the other adjustment perforations toward the outside or in the direction of the same toward the inside and surround the respective 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 about the narrow region and the other aligning foot is in a position to engage the other adjustment perforation/perforations in this position.

The walls of the enlarged adjustment perforations should have the same configuration on the side facing toward or away from the other adjustment perforations as the aligning feet to be inserted. As a result, when centered in the exactly defined positions in which they would also be inserted, as is known, circular. The adjustment perforations not made of a common diameter, namely one of about 5-6 mm. The enlarged adjustment perforations nearly their halves the same configuration from where they would reach a larger transverse dimension of about 9-10 mm. The enlarged adjustment perforations open toward the underside of the sound carrier. Actually, two adjustment perforations which one is circular and the other funnel-shaped are sufficient. Nevertheless, mostly four perforations are provided, as has been known, of 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 contributes to 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 of about 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 an annular bead does not only make the perforations better visible, they also render the alignment of the feet of the sound reproducing device easier because they practically fall into the larger annular bead and slip into the precise adjustment perforations automatically	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 dislocation of 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 for interruptions 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 Lernertexten. 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 away from the other adjustment perforations toward the outside or in the direction of the same toward the inside and surround the respective 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 about the narrow region and the other aligning foot is in a position to engage the 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 in sie einzuführenden Justierfüss mit dem so geschaffenen breiten Bereich mit grossem Spiel umschliessen, dass dieser Justierfüss 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 passes the sound reproducing device placed in position into the round perforations of the sound carrier	1-1	99.86%	So ist ein ebener Führungsrichter 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 from the 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 case in the exactly defined			Dadurch befinden sich die Füße lagestabil in der zentrierten Lage jeweils an den genau

ist. Dieser Wulstring kann sich unmittelbar lässt 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

urchrüche am unteren Rand des Tonträgers mit diesem Wulstring ausgestattet. 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. Man kann diese von einem Wulstring umgebenen Durchbrüche praktisch mit einem Trichter vergleichen.

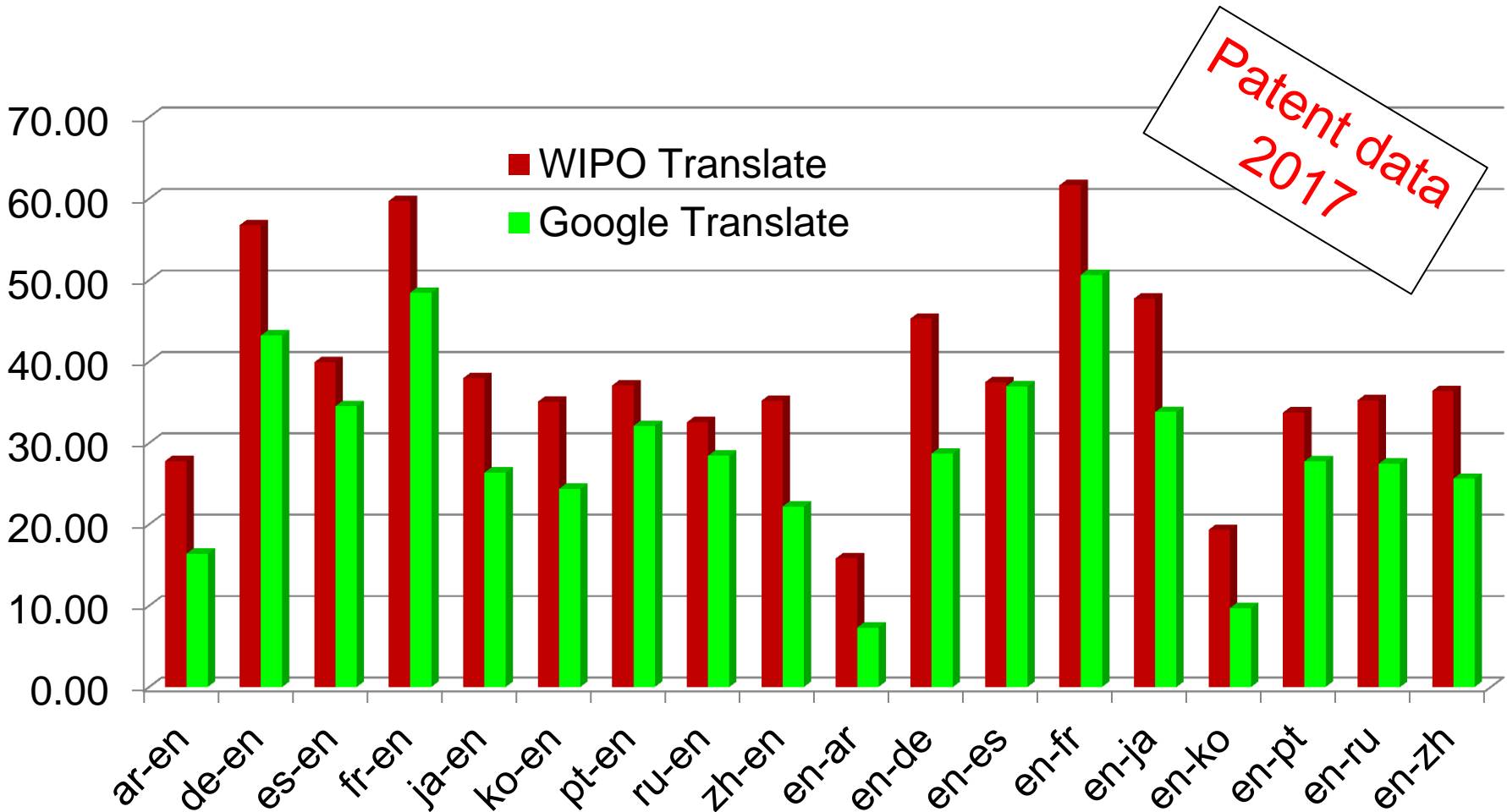
Durchbrechungen eines Tonträgers mit Wulsten versehen werden. Diese Wülste verhindern auch das unbeabsichtigte Verrücken des Tonwiedergabegerätes bei evtl. gewollten Unterbrechungen der Wiedergabe beim Studium von Lernertexten. Das ist wichtig, weil das Tonwiedergabegerät auch während der Wiedergabe häufig, beispielsweise für Unterbrechungen oder Wiedergabewiederholungen zu betätigen ist und dabei an sich die Gefahr des Verrückens besteht.

Alternativ zu dieser Lösung kann auch ein Teil der Justierdurchbrechungen, z. B. zwei untere Justierdurchbrechungen, in Richtung von den anderen Justierdurchbrechungen nach aussen hin verbreitert und verlängert ausgebildet sein und in sie einzuführenden Justierfüss mit dem so geschaffenen breiten Bereich mit grossem Spiel umschliessen, dass dieser Justierfüss 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. So ist ein ebener Führungsrichter gebildet, der das aufgesetzte Tonwiedergabegerät in die runden Durchbrechungen des Tonträgers leitet

Zusammen mit den dann zum Eingriff kommenden anderen Justierfüssen ist das Tonwiedergabegerät dann lagestabil gehalten. 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

Ausserdem verdeckt das Tonwiedergabegerät jegliche Sicht zur schuhartigen Vertiefung. Die Wände der vergrösserten Justierdurchbrechungen sollten auf der den anderen Justierdurchbrechungen zugewandten oder abgewandten Seite die gleiche Gestalt haben wie die einzusetzenden Justierfüsse. Dadurch befinden sich die Füße lagestabil in der zentrierten Lage jeweils an den genau definierten Positionen, an denen die Justierfüsse einzugreifen vermögen. So ist ein ebener Führungsrichter gebildet, der das aufgesetzte Tonwiedergabegerät in die runden Durchbrechungen des Tonträgers leitet

Quality – all languages



BLEU score comparison between WIPO Translate and Google Translate (both using NMT models), testset containing titles and abstracts from patents published after July 2017(except Arabic). Tested uniquely with new sentences NOT used in the training of WIPO Translate

WIPO Translate: in different places



NMT project for 2018?



On-going:



Expressed interest:

NMT planned 2018?



Typical collaboration WIPO Translate for IPOs

User agreement: WIPO Translate for free, but WIPO keeps intellectual property protection

Hardware: IPO needs to buy a GPU server (see technical documentation)

Software: IPO gets all the software (and source code, and updates).

Configuration: DIY or provide remote access to WIPO engineer

Transfer knowledge: 100 page documentation, WIPO engineer can be invited

Operations: IPO needs to take over the maintenance of this server

Conclusion and discussion

- ATAC aims at using in-house development & open source
- In domain data creates usually better tools
 - Test it with in-domain data only 😊
 - WIPO translate trained on patents does not know how to translate "Who am I?" 😊
- Future work:
 - Offer non-English language pairs (e.g. French→Chinese, Arabic→Russian...)
 - Explore many-to-many models (e.g. one single model to translate from English/French/German//Portuguese /Spanish into Chinese/Japanese/Korean)
 - Explore NMT for transliteration



Other applications of (supervised) machine learning

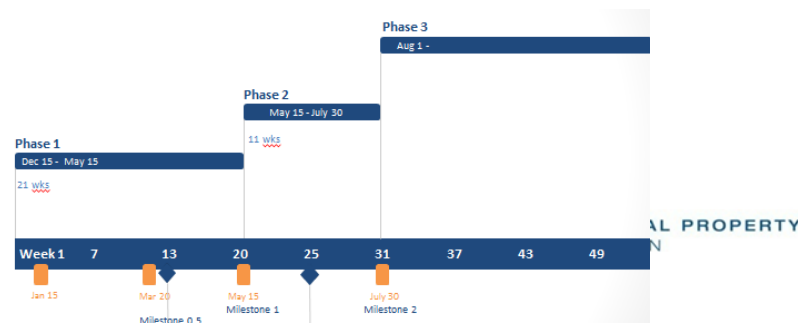
Bruno Pouliquen,
World Intellectual Property Organization (WIPO)

May/2018

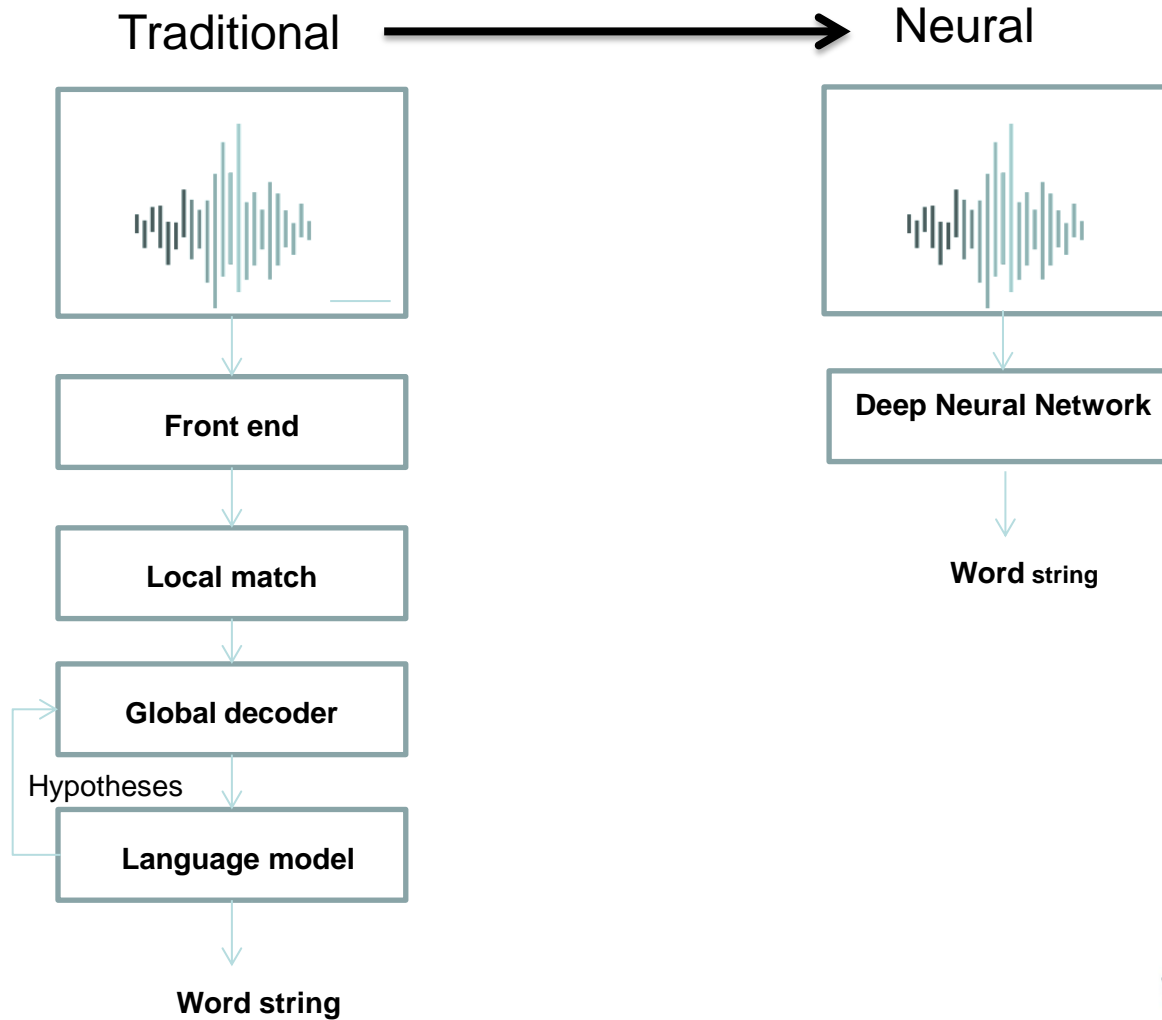
Speech to text project

- Use records of WIPO meetings and corresponding transcripts
- Use machine learning techniques to output corresponding text
- Use cases:
 - Captioning
 - Send verbatim for translation
 - Text based search engine
 - ...

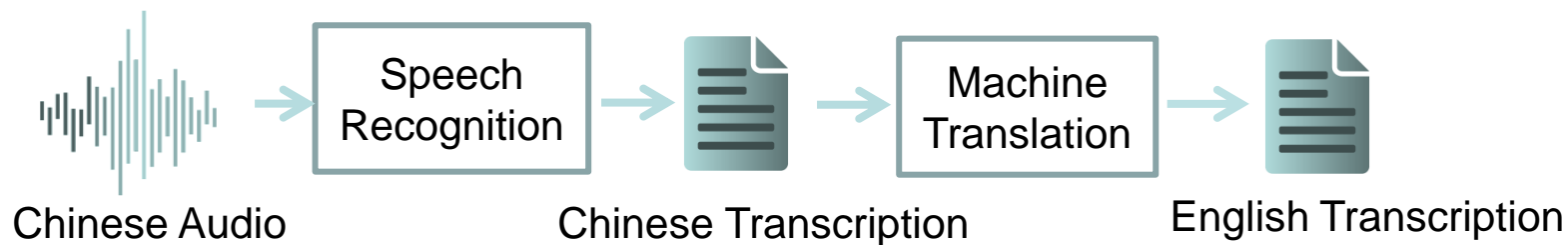
Exploratory project: currently collecting data (early prototype ~ September)



Speech to text project



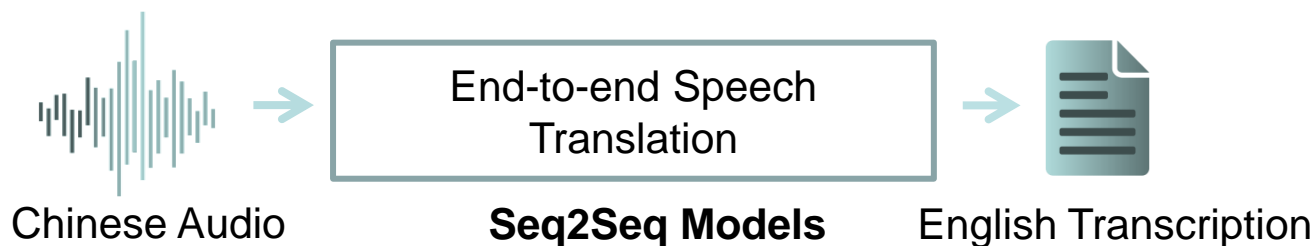
Speech to translated text



- Drawback:

- Accuracy
- Latency
- Low resource settings

⇒ **Do it DIRECTLY**



Early research project: will not go in production

Text based classification

- “learn” from pre-classified patent applications
 - e.g. title and abstract => IPC code
- “guess” the best classes for a new patent application
 - e.g. output IPC code (& score) from a new title & abstract

Proper names: classification

- “guess” country of a person name:
 - HAN, Guydon => KR
 - CHEN, Lili => CN
 - Viswanathan, Anand=>IN
 - Wojtaszek, Radoslaw=>PL
 - ...
- Proper name classification (company or person?):
 - Metal Paris => Company
 - Paris Overton => Person

Demo

Proper name transliteration



Demo

- Similar to machine translation
 - “translation” from/to different scripts
- “learn” from existing transliterations:
- “guess” possible spelling in different script
 - LI, Lili => 李丽丽
 - Francis Gurry=> 弗朗西斯·古里
 - 埃米·安·瓦兰斯基 => Emy Ann Walanski
 - 可口可乐 => Coca-Cola

Applicants: Boditechmed Inc.
韩国帕克特生物科技有限公司

Inventors: Choi Euiyeol
崔义烈
Nam Kibong
南基凤
Kim Jaehoon
金载勳
Jung Dongseok
郑东锡
Park Sangyeol
朴相烈
Moon Jungdae
文政大
Jung Jinha
郑晋河
Kim Youngmin
金永敏
Jung Soyoung
郑素映
Park Aekyung
朴爱京
Kim Byoungchul
金丙澈
Kim Sungjoong
金成中

Image classification...