Name standardization

(experiments with name transliteration in WIPO)

2 May 2019

Bruno Pouliquen
Global Database Division

Introduction

"Advanced Technology Applications Centre"

Explore machine learning techniques in the IP domain

- Text:
  - Neural Machine Translation
  - Automatic classification
- Speech
- Image
- …

Is transliteration a simple task?

… does it impact mame standardization?

- معمر القذافي(ar,ps)
- Муамар Каддафи (bg,ru)
- Mouammar Kadhafi (Eu,rw)
- Muammar Gaddafi (Eu,sw)
- Moammar Gadhafi (da,sw)
- Muammar al-Gaddafi (Eu,yo)
- Muammar Kadhafi (en,ro)
- معمر قذافی(fa)
- Моамар Кадафи (bg)
- معمر القذافی(fa,ur)
- Муамар Кадафи (bg)
- Muammar Gheddafi (it)
- …

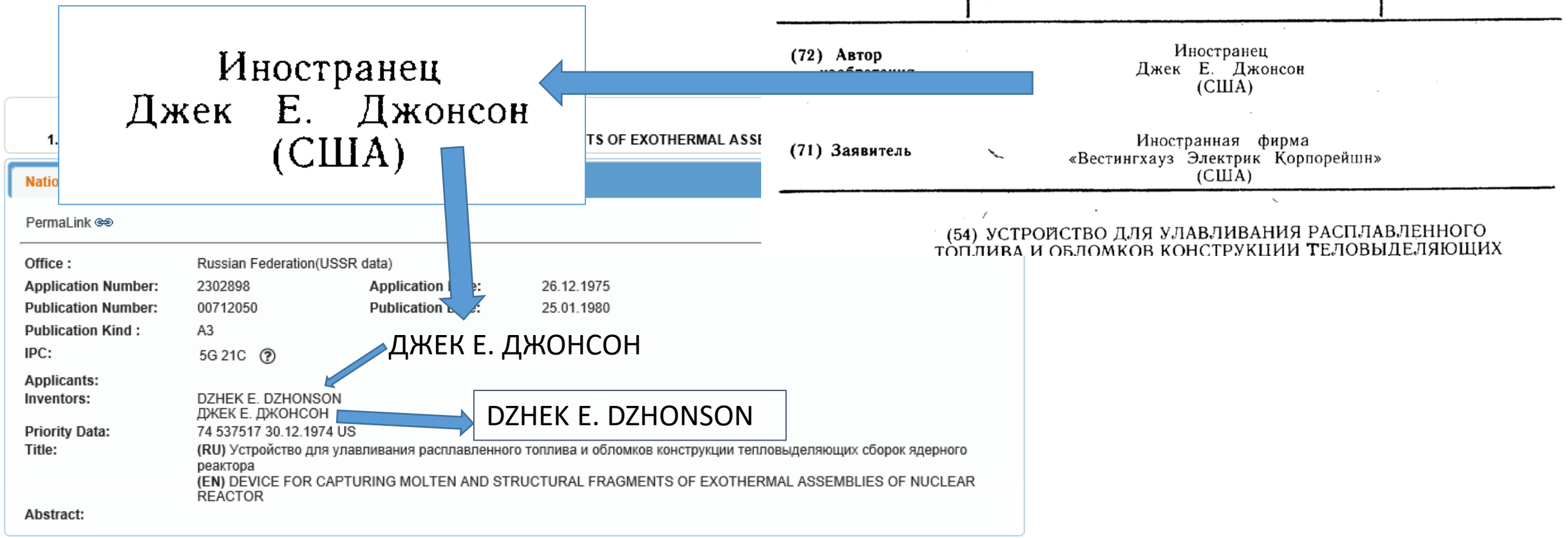
100 orthographies ;-)

معمر القذافي(ar,ps)
Муамар Каддафи (bg,ru)
Mouammar Kadhafi (Eu,rw)
Muammar Gaddafi (Eu,sw)
Moammar Gadhafi (da,sw)
Muammar al-Gaddafi (Eu,yo)
Muammar Kadhafi (en,ro)
معمر قذافی(fa)
Моамар Кадафи (bg)
معمر القذافی(fa,ur)
Муамар Кадафи (bg)
Muammar Gheddafi (it)
Muammar Kaddafi (da,tr)
Muammer Kaddafi (da,tr)
Muamar Gadafi (es)
Moamerja Gadafija (sl)
Moamar Gaddafi (da,sw)
Muamar el Gadafi (en,pt)
Muamar Kadafi (eo,pt)
Muammar Kadafi (en,pt)
Muammar Khadaffi (es,sv)
Moammar Gaddafi (da,sw)
Moamer Kadhafi (en,nl)
Moammar Kadhafi (fr,sv)
Muammar Gadafi (en,sw)
Muammar el Gaddafi (Eu,es)
Moamer Gadafi (sh,sl)
Muammar Gadaffi (da,sw)
Mouammar Khadafi (fr,pt)
Муаммар Каддафи (os,ru)
Moammer Kadhafi (en,nl)
Muammar Qaddafi (en,sw)
Muammar Khadafi (en,ro)
Muamar Kadhafi (es,pt)
Muammar al Gaddafi (de,hr)
Muamar al Gadafi (en,pt)
Muammar Khaddafi (nl,sv)
Muhammar Gheddafi (it)
Moamer Gathafi (en)
Muammar el-Qaddafi (en,pt)
Muammar Gadhafi (de,sw)
Moammar Gadaffi (da,sv)
Mouammar Kaddafi (fr,nl)
Muammar Kadaffi (nl,sv)
Muamar Khadafi (es,pt)
Muammar al-Ghadhafi (de)
Muammar Gadafy (en)
Muamar Gaddafi (da,sw)
Muammar al-Gadafi (es,no)
Moamer Gaddafi (da,sw)
Moammar al-Kadhafi (nl)
Moammar Qaddafi (en)
Muammar al-Qadhafi (de,sw)
Muammar Ghadhafi (de)
Moammar Kadafi (en,pt)
Muammar Qadhafi (de,sw)
Moamar Gadafi (es,sl)
Muammer Gaddafi (da,sw)
Mouammar El Kadhafi (fr)
Mouamar Kadhafi (fr,pt)
Muammer Gadaffi (Eu,en)
Muhammar Gaddafi (de,sv)
Moammar Gheddafi (it)
Muammar al-Qaddafi (Eu,jv)
Muammar al-Kadhafi (fr,nl)
Mummar Gaddafi (en,no)
Muammar Gaddaffi (en,sv)
Moammar Ghadafi (en)
Moammar al-Qadhafi (de,nl)
Moammar Khadaffi (nl,sv)
Mouamar Khadafi (en,pt)
Moamar Kadhafi (en,pt)
Moamer Gadhafi (en)
Muammar Ghadaffi (en,sv)
Moammar Khadafi (fr,pt)
Muammar Al Gaddafi (de,ms)
Moeammar Kadhafi (nl)
Moamer Kadaffi (nl)
Moamer Qadhafi (en)
Muammar el-Gaddafi (de,en)
Muammar al-Qathafi (en)
Muammar Al-Qadhafi (en,sw)
Muammar Ghadafi (en,sw)
Moammer Gaddafi (en,sw)
Moammer Gadhafi (en)
Muamar El Gadafi (es)
Muammar Al-Gaddafi (de,no)
Muammar el Gadafi (es)
Muamerja Gadafija (sl)
Moammar Kaddafi (fr,pt)
Muhamar Kadhafi (pt)
Muammar al-Gadhafi (de,ro)
Mouammar Al Kadhafi (fr)
Muammar Khadafy (es,pt)
Moammar Khaddafi (nl,sv)
Moammar Khaddafy (en)
Muammar Kadhaffi (pt)
Muammar Al-Kaddafi (pl)
Moammar Qadhafi (de,en)

Source: http://emm.newsexplorer.eu/NewsExplorer/entities/fr/262.html

Transliteration: challenges

- E.g. application SU27926309



Transliteration, back-transliteration

- Transliteration:
  - Jack E. Johnson ➔ДЖЕК Е. ДЖОНСОН
- Back-transliteration:
  - (good) ДЖЕК Е. ДЖОНСОН ➔ Jack E. Johnson
  - (bad) ДЖЕК Е. ДЖОНСОН ➔ DZHEK E. DZHONSON

Future/current work on transliteration

already on Patentscope (currently only for Japanese)

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
aAtao Leonard F

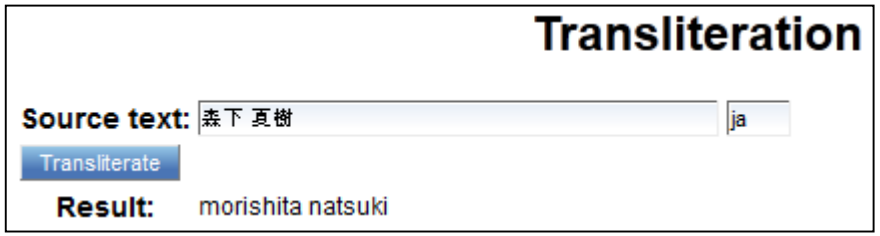
Shusaku Yamamoto
Yasumura Takaaki
Natsuki Morishita

Name transliteration

- Character based sequence-to-sequence monotonic decoding
- NMT “learns” transliteration rules from “noisy” examples

Аанестад Лейф **Инге**	Aanestad Leif I. (Inge)
ААН ХЕНДРИК КЛАЗИНГА	**KLASINGA** AAN HENDRIK
АГРЕ **Дэниел** Х.	D. H. Agre
Альбрехт Эденхофер	**dr** Albrecht Edenhofer

Prototype:



فيليب ماجنير — Philippe Magner

карлос абад — Carlos Abad

鲍里斯·布里亚肖夫 — Boris Blyashov

Challenges:

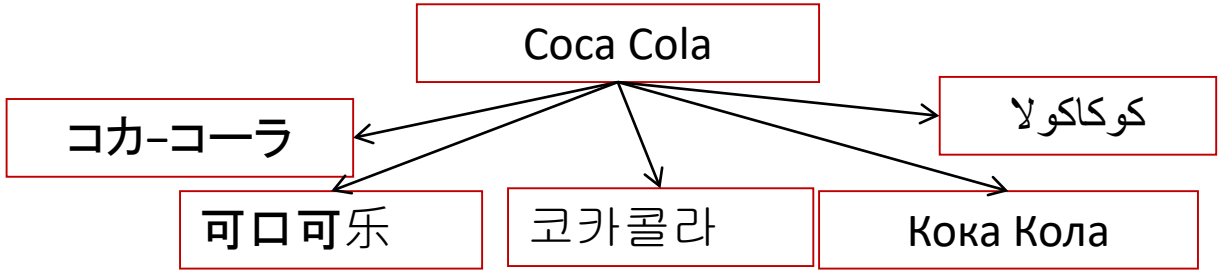
Хавьер → Javier (Luis **Javier** Juarez Vargas)

Хавьер → Xavier (**Xavier** Chatellard)

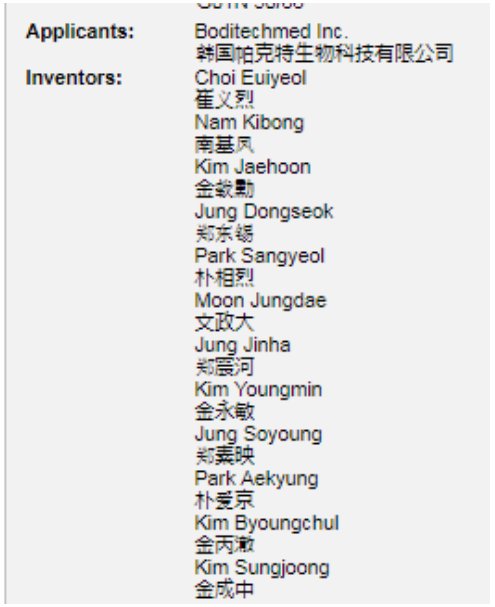
Arthur → Артюр (Jacques Marcel **Arthur** Bunel)

Arthur → Артур (**Arthur** Mark Thomas)

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
  - 可口可乐 => Coca-Cola

Other examples on using AI on proper names

- “learn” from patent applicant/country
- “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

Early prototype available (English only)

Applicant name standardization (future work)

- Use machine learning
- “learn” from original-name/standardized-name
- “guess” standardized version of a new name
  - Apple => APPLE, INC.
  - Applle, inc => APPLE, INC.
  - …

- Get access to training data
- Machine learning alone? Hybrid (combined with rules)? human supervision?

Discussion/future work

- This is exploratory work ;-)
- Future includes:
  - Build complete transliteration models:
    - Latin
    - Korean (Hangul)
    - Chinese
    - Japanese (Katakana+Hiragana+Kanji)
    - Arabic
    - Cyrillic
  - Integrate transliteration on Patentscope
  - Gather more/cleaner training examples
  - Address the Neural Machine Translation specific problems with proper names