



MT by EUCTED Tilde DE-PL

100 MILIONÓW SŁÓW PRZETŁUMACZONYCH TRANSLATOREM EU COUNCIL PRESIDENCY!

19.11.2020 | język i rozumienie tekstów | technologia językowa i wielojęzyczność | Saarbrücken

www.presidencymt.eu

Powered by DeepL DFKI Connecting Europe eTranslation TILDE

Uwaga: nieco skorygowana polska wersja wywiadu niemieckojęzycznego przygotowana z EUCPT (Tilde Engine).

Wywiad z prof. dr. Josefem van Genabith, kierownikiem działu badań nad wielojęzycznymi technologiami językowymi w DFKI w Saarbrücken, na temat tłumaczeń maszynowych i translatora EU Council Presidency, który jest wykorzystywany w ramach niemieckiej prezydencji w radzie UE od lipca 2020 r.

- Panie profeszorko van Genabith, jest pan dyrektorem naukowym DFKI i od 2014 roku kieruje badaniami nad wielojęzycznymi technologiami językowymi (MLT) w Saarbrücken. Jak wyglądała pana kariera naukowa przed przeniesieniem się do Saarbrücken?

Sukces translatora EU Council Presidency to piękna okazja dla naszego teamu MLT i naszych partnerów w [DeepL](#), [Tilde](#) i e-[Translation](#)! Jestem bardzo dumny z zespołów i pracy, którą wykonali w ścisłej współpracy z Federalnym Urzędem Spraw Zagranicznych. Osobiście od dawna interesowałem się językiem i technologią, studiowałem elektrotechnikę i anglistykę w RWTH Aachen, a następnie miałem szczęście: dzięki stypendium Rady brytyjskiej, a następnie Biura Spraw Zagranicznych i Wspólnoty Narodów, na Uniwersytecie w Essex uzyskałem najpierw tytuł magistra, a następnie zrobiłem doktorat u Louisy Sadler. Na początku lat 90. byłem na studiach podoktoranckich u Hansa Kampa z Instytutu Przetwarzania Mowy Maszynowej (IMS) w Stuttgarcie. Świetny czas! Następnie spędziłem 17 lat w Irlandii na wydziale informatyki na uniwersytecie w Dublinie, gdzie przeszedłem całą procedurę od wykładowcy, przez starszego wykładowcę, aż do pozycji profesora nadzwyczajnego. W Dublinie miałem wiele swobód i wspaniałych kolegów z DCU, z innych uniwersytetów w Dublinie i z licznych firm z branży zaawansowanych technologii z siedzibą w Irlandii

(IBM, Microsoft, Symantec) i mogliśmy korzystać z tych swobód: odnowiłem Narodowe Centrum Technologii Językowych (NCLT) i byłem dyrektorem założycielem CNGL (Centrum Lokalizacji Nowej Generacji, obecnie ADAPT kierowane przez Vinny'ego Wade'a). Dzięki tym pracom, a w szczególności CNGL, na początku drugiej połowy lat 2000-2010 coraz bardziej angażowaliśmy się w międzynarodowe projekty, takie jak projekty unijne, w których byłem dyrektorem naszego laboratorium w Saarbrücken, Hans Uszkoreit, był bardzo aktywny. Hans Uszkoreit, który w międzyczasie zbudował siostrzane laboratorium w Berlinie (obecnie SLT, kierowane przez Sebastiana Möllera), przybył do Saarbrücken i DFKI w Irlandii w 2014 r. po 17 latach.

- Oprócz pracy w DFKI, piastuje pan również katedrę na uniwersytecie w Saarze. W jaki sposób praca akademicka i praca praktyczna wzajemnie się uzupełniają?

Najważniejszym elementem naszej pracy są nasi pracownicy: dzięki nim nasza praca odniesie sukces. Moi współpracownicy z uczelni i DFKI współpracują ze sobą w kolorowej mieszance zespołów. Na cotygodniowych wspólnych spotkaniach nie ma różnicy, czy ktoś jest w DFKI, czy na uniwersytecie. Jesteśmy częścią [SFB1102](#) (Gęstość informacji i kodowanie językowe) na uniwersytecie, mamy projekt DFG na uniwersytecie dotyczący multimodalnej post-edycji, w którym z powodzeniem współpracujemy z zespołem DFKI profesora Antoniego Krügersa; kieruję europejskim programem magisterskim z technologii językowej i komunikacyjnej ([LCT](#), Erasmus +), który jest doskonale zarządzany przez jedną z moich współpracowniczek pracujących w laboratorium [MLT](#) (DFKI) na niepełnym etacie. Wszyscy moi pracownicy zarządzający DFKI w czterech grupach MLT: *Tłumaczenie maszynowe*, *Odpowiadanie na pytania i Ekstrakcja informacji*, *Mówiące roboty* oraz *Dane i zasoby ucza*, prowadzą seminaria i szkolą studentów studiów doktoranckich, magisterskich i licencjackich. Podobnie wiele współpracowników z zespołów MLT pracuje na uniwersytecie. Oczywiście pod względem formalnym i finansowym wszystko jest w czysty sposób rozdzielone na projekty. Ale związek z uczelnią jest bardzo silny. Wydział Nauki i Technologii Językowych na uniwersytecie w Saarze jest jednym z najlepszych w Europie. W laboratorium MLT na DFKI prowadzone są bardzo intensywne badania: na przykład w 2020 r. opublikowaliśmy ponad 10 dokumentów na głównych międzynarodowych konferencjach w tej dziedzinie (ACL, ICML, EMNLP, COLING, IJCAI) poświęconych technologiom językowym, sztucznej inteligencji i uczeniu maszynowemu. To wielki sukces i pokazuje jakość zespołów. Z drugiej strony, ukierunkowane na zastosowanie badania DFKI stanowią atrakcję dla studentów i naukowców z uniwersytetu: gdzie indziej, tak jak na przykład w EU Council Presidency Translator, ich praca jest publicznie wykorzystywana w taki sposób, że 100 milionów słów zostało przetłumaczonych w ciągu 4,5 miesiąca (do dziś)? To naprawdę świetne!

- EU Council Presidency Translator jeszcze bardziej zwiększył w Niemczech widoczność usług w zakresie tłumaczeń maszynowych. Jest to wspólne działanie wielu podmiotów, ale pan kierował tym projektem. Kiedy zaczął pan tę pracę? Jak zorganizował pan konsorcjum? Ilu naukowców było w to zaangażowanych?

[EU Council Presidency Translator](#) jest bardzo europejskim rozwiązaniem, które pokazuje, że Europa jest bardziej niż konkurencyjna na arenie międzynarodowej w dziedzinie technologii językowych i sztucznej inteligencji na najwyższym poziomie: opiera się na połączeniu doskonałej wiedzy specjalistycznej w dziedzinie zaawansowanych technologii i sztucznej inteligencji w Niemczech (DeepL, DFKI), na Łotwie (Tilde) i Komisji Europejskiej (e-Translation). To partnerstwo między przemysłem (DeepL, Tilde), władzami publicznymi (EC, eTranslation) i instytucją badawczą (DFKI). DFKI kieruje projektem, a dofinansowanie pochodzi z Federalnego Ministerstwa Spraw Zagranicznych, które sprawuje przewodnictwo w niemieckiej prezydencji Rady UE. Kompetencje

członków konsorcjum doskonale się uzupełniają: przez wiele lat Tilde, korzystając ze wsparcia europejskiego, opracowała podstawową strukturę tłumacza Presidency, w który wbudowane są maszyny tłumaczeniowe wielu dostawców. DeepL oferuje wysokiej jakości maszynę tłumaczeniową dla 8 języków. e-Translation (KE) zapewnia podstawowe usługi tłumaczeniowe dla wszystkich 24 języków urzędowych UE. W ścisłej współpracy ze służbami tłumaczeniowymi ministerstw DFKI opracował systemy automatycznego tłumaczenia na język niemiecki, francuski i hiszpański, dostosowane specjalnie do danych i potrzeb ministerstw. Tilde robi to w językach angielskim, włoskim i polskim. W DFKI Stephan Busemann zapewnia obsługę administracyjną tłumacza Presidency. Ja kieruję aspektami naukowymi i technicznymi. Cristina España Bonet, kierowniczka zespołu MT w laboratorium MLT i jej współpracownicy, Jingyi Zhang, opracowują systemy. Pomagają im dwie studentki, Damjana Gateva i Anastasija Amman, z programu magisterskiego „Nauka o języku i technologia”. DFKI prowadzi również działania informacyjne i medialne tłumacza Presidency. Jest to nadzorowane przez Eileen Schein i jej koleżankę, Marlies Thönnissen, w zespole MLT i aktywnie wspierane przez Departament Komunikacji Korporacyjnej DFKI.

- Wykorzystują państwo sztuczne sieci neuronalne do tłumaczenia. Proszę opisać, jak działa państwa maszyna tłumaczeniowa.

Modele neuronalne umożliwiły w ostatnich latach skoki kwantowe w jakości wielu technologii językowych i innych zastosowań w sztucznej inteligencji. Nasze systemy wykorzystują głębokie sieci neuronowe oparte na modelach transformatorów. Modele te wykorzystują różne rodzaje uwagi i są w dużej mierze wysoce zbieżne.

- Sztuczne sieci neuronowe są testowane z bardzo dużą ilością danych głosowych. Skąd pochodzą te dane treningowe i testowe i w formie szacunkowej, ile to słów w toku?

Dla wielu par językowych nasze dane treningowe obejmują dziesiątki milionów par zdań, przy czym każda para zdań zawiera punkt wyjścia w jednym języku i jego tłumaczenie na inny język. Na tej podstawie maszyny uczą się tłumaczyć. Dane te opierają się na tłumaczeniach już wykonanych przez ludzi. Maszyna uczy się od ludzi. Dane pochodzą ze zbiorów danych UE, [ELRC](#) (Koordynacja europejskich zasobów językowych, którą kierujemy również w MLT DFKI) i innych źródeł. Ponadto ściśle współpracujemy z zespołami tłumaczeniowymi ministerstw, aby wykorzystywać dane z ministerstw do tworzenia specjalnych maszyn, które są szczególnie dostosowane do potrzeb ministerstw. Są one stale oceniane przez tłumaczy z ministerstw, co umożliwia ich ciągłą poprawę w trakcie realizacji projektu.

- Tłumacz Presidency był intensywnie wykorzystywany przez użytkowników w ciągu ostatnich 150 dni. Przetłumaczono ponad 100 milionów słów. Jakie były najbardziej poszukiwane pary językowe? Czy były też zdania, które szczególnie często występowały?

W przeciwieństwie do innych ofert tłumacz Presidency jest pewny i bezpieczny, wszystkie serwery są w UE, transmisje są zaszyfrowane, a po wykonaniu tłumaczenia wszystkie dane są natychmiast usuwane. Mamy więc do wykorzystania tylko informacje wysokiego poziomu. Dane liczbowe pokazują, że tłumaczenie niemieckojęzycznej strony internetowej prezydencji Rady „za jednym kliknięciem” zostało bardzo dobrze przyjęte: daje to około 47% ze 100 milionów przetłumaczonych do tej pory słów. Preferowanymi językami docelowymi tłumaczenia maszynowego na stronie internetowej prezydencji Rady są: hiszpański, włoski i portugalski (wersje francuska i angielska zostały

sporządzone ręcznie). Nieco większa połowa wynika z tłumaczeń tekstowych (22%), pisemnych (30%) i stron internetowych (2%) na stronie tłumacza, przy czym najbardziej potrzebne jest tłumaczenie na język niemiecki i angielski.

- Co mówią tłumacze o nowej jakości tłumaczenia maszynowego? Czy tłumacze uważają maszyny za konkurencję czy za narzędzia wspierające ich pracę? Jak zmienia się profil zawodowy tłumacza?

W ramach projektu „EU Council Presidency Translator” bardzo ściśle współpracujemy z kolegami ze służb tłumaczeniowych ministerstw: zajmują się gromadzeniem i udostępnianiem danych w ministerstwach w celu dostosowania specjalistycznych maszyn do potrzeb ministerstw. Ponadto badają i oceniają specjalistyczne maszyny i dzięki swoim wynikom przyczyniają się w sposób scentralizowany do doskonalenia systemów. W procesie tłumaczenia maszyna jest narzędziem pomocniczym: w przypadku dobrej jakości tłumaczenia maszyna może pomóc w zwiększeniu wydajności tłumacza. Zmienia to wizerunek zawodowy tłumacza w kierunku kontroli jakości, zapewniania jakości poprzez redagowanie (poprawianie) tłumaczeń automatycznych oraz certyfikację i jakość tłumaczeń. Współczesne szkolenia tłumaczy uwzględniają te zmiany: studia tłumaczeniowe „Translatoryka i technologia” na Uniwersytecie w Saarze mają komponent zaawansowany technologicznie, w ramach którego przyszli tłumacze są zaznajomieni z technologiami językowymi opracowywanymi przez ich kolegów ze studiów lingwistyki komputerowej (Translatoryka i technologia) i informatyki.

- Niemiecka prezydencja Rady UE kończy się 31.12.2020. W jaki sposób następnie wykorzystywany będzie translator Presidency? I niezależnie od tego, jakie są pańskie dalsze plany?

Translator Presidency został bardzo dobrze przyjęty i przewyższył wszystkie dotychczasowe rekordy poprzedniego translatora Presidency. Jestem bardzo dumny z tego, co zespół MLT w DFKI wykonał wspólnie z kolegami z DeepL, Tilde i e-Translation! Istnieje duże zainteresowanie wykorzystaniem translatora Presidency na kolejnych prezydencjach Rady. Trwają rozmowy na ten temat. Branża jest również żywo zainteresowana niemiecką i europejską technologią językową: technologią językową i sztuczną inteligencją „made in Europe”. Tłumaczenie maszynowe jest tylko jedną z kompetencji w naszym podręczniku MLT: inne to grupa „Pytań i odpowiedzi i Ekstrakcji informacji” (w szczególności w dziedzinie biomedycyny), grupa „Mówiące roboty” (skupiająca się na systemach dialogu i robotyce ratunkowej) oraz grupa „Dane i zasoby” (która od wielu lat kieruje dużymi projektami UE, takimi jak ELRC). Do tego dochodzi nasza siostrzana [technologia SLT](#) (Technologia mowy i języka) w Berlinie. Obydwa laboratoria (MLT w Saarbrücken i SLT w Berlinie) ściśle ze sobą współpracują i uzupełniają swoją wiedzę fachową.