

Schlussbericht gefördertes Forschungsprojekt

Projekt Titel: Translatur-ia (Translaziun Rumantscha cun intelligenza artificizala)
Abteilung/ Fachbereich FHGR: Schweizerisches Insitut für Informationswissenschaft

Antrag bewilligt: Dezember 2019
Gesprochener Förderbetrag: CHF 40'000

Projektleitung/ Projektverantwortung:
Kuntschik, Philipp, +41 81 286 37 35, philipp.kuntschik@fhgr.ch

Projektteam:
Kuntschik, Philipp, SII
Rölke, Heiko, SII
Kalbermatter, Urban, SII
Schmid, Marco, SII
Peers, Lisa, ITF

Kooperationspartner: Lia Rumantscha, Standeskanzlei Graubünden
Geplante Laufzeit des unterstützten Projektes: 01/2020 - 12/2020

1) Management Summary:

Translatur-ia beabsichtigt Anwendungen, welche im Zusammenhang mit der Rätoromanischen Sprache stehen (zum Beispiel Übersetzungsdienstleistungen), durch die Entwicklung und Einbindung von Computertechnologie zu unterstützen. Im Projekt zum vorliegenden Schlussbericht wird aufgezeigt, dass die Schaffung solcher Technologien durchaus realistisch ist. Konkret wurde aufgezeigt, wie sich der Übersetzungsprozess durch die Einbindung einer prototypischen Übersetzungshilfe in Form eines Web-Dienstes effizienter gestalten liesse.

a) Projekt Ziele laut Antrag sowie geplantes/r Vorgehen/ Ansatz:

Ziel des Projekts war es, ein Werkzeug zu schaffen, sowie Empfehlungen auszuarbeiten, um Unternehmen mit Übersetzungstätigkeit dabei zu unterstützen, ihre bestehenden Prozesse besser zu digitalisieren und teilweise zu automatisieren. Durch diese Ziele sollte entsprechende Fachkompetenzen an der FHGR aufgebaut, und ein Unterstützungsnetzwerk gebildet werden, welches es erlaubt das geschaffene Werkzeug in Zukunft zu optimieren und damit zu einem Gegenstand des täglichen Gebrauchs weiterentwickeln zu können.

b) Konkrete Ergebnisse/ Resultate:

c) Das geschaffene Werkzeug kann unter der Web-Adresse <https://translaturia.fhgr.ch> aufgerufen und genutzt werden. Es wurde somit aufgezeigt, dass eine nutzbare Übersetzungshilfe trotz der äusserst geringen verfügbaren Datenmenge denkbar und umsetzbar ist. Eine deutliche Steigerung der Ergebnisqualität ist durch eine massive Erhöhung der Datengrundlage erreichbar. Das Forschungsteam konnte die notwendige Kompetenz entwickeln um diese Verbesserung in zukünftiger Forschungsarbeit zu verfolgen und auf vergleichbare Situationen anwenden zu können.

2) Detaillierte Projektbeschreibung:

a) Projektteam:

Kuntschik, Philipp, FHGR, Projektleitung, Forschung & Entwicklung
Rölke, Heiko, FHGR, Controlling & Forschung
Kalbermatter, Urban, FHGR, Forschung & Entwicklung
Schmid, Marco, FHGR, Forschung & Deployment
Peers, Lisa, FHGR, Qualitätssicherung & Datenakquise

b) Kooperationspartner:

Klaiss, Conradin, LR, Datenakquise & Qualitätssicherung
Telli, Daniel, LR, Qualitätssicherung
Dermont, Beat, Staka, Datenakquise

c) Projekthintergrund:

Das Bundesamt für Kultur sieht laut einem im Auftrag erstellten Evaluationsbericht (2019) mittelfristig die Gefahr einer existentiellen Bedrohung der rätoromanischen Sprache. Es geht hervor, dass unter anderem die mediale Präsenz fehlt um der Sprache die benötigte Sichtbarkeit zu geben. Das Projekt versucht durch die Teilautomatisierung des Übersetzungsprozesses Potential zur Steigerung von Qualität, Geschwindigkeit und Effizienz von Übersetzungsaufgaben zu schaffen.

Computergestützte Übersetzungen sind ein Anwendungsfeld der Computerlinguistik, und erzielten mit dem Aufkommen von Deep Learning basierten Algorithmen in den letzten Jahren bei grossflächig gesprochenen Sprachen fast menschenähnliche Qualitätsniveaus. Bei diesem neuartigen Ansatz werden immense Datenmengen genutzt, um ein Computersystem durch iterative Trainings zu schaffen. Im Fall der Rätoromanischen Sprache existiert eine nur bedingt ausreichende Menge an Daten, die konkrete Disziplin wird deshalb «Low Resource Machine Translation» bezeichnet, aktuelle Ergebnisse liegen hierbei weit hinter vergleichbaren stark vertretenen Sprachen zurück. Ansätze zur Verbesserung der Übersetzungsqualität in dieser Disziplin versuchen die verfügbare Datenmenge künstlich zu erhöhen, sowie mit der verfügbaren Datenmenge effizienter umzugehen.

d) Projektziele:

Zu den Projektzielen zählt die i) Entwicklung von Technologien zur Unterstützung von Übersetzungstätigkeiten in Form eines Computerprogramms, sowie die ii) Ausarbeitung von Empfehlungen zur Implementierung dieses in die Übersetzungstätigkeit. Durch diese Arbeit soll der iii) Kompetenzaufbau im Bereich Machine Translation und die iv) Entwicklung eines Kooperationsnetzwerks gefördert werden. Es soll somit die Ausgangslage geschaffen werden, um im Bereich Machine Translation national konkurrenzfähig zu sein.

e) Projektverlauf/ Vorgehen/ Termine:

Das Projekt konnte wie geplant durchgeführt werden, manche Arbeitspakete erwiesen sich als langwieriger als geplant: So muss das Arbeitspaket 3 «Datenakquisition» als nicht abschliessbare Daueraufgabe betrachtet werden. Zwar konnten bis zum Meilenstein 2 «Korpus Auswahl abgeschlossen» ausreichend Daten für ein erstes Training bereitgestellt werden, allerdings zeigte der Prozess auf, dass das System direkt von einer grösseren verfügbaren Datenmenge profitiert. Der Korpus wird deshalb über die gesamte Projektdauer und darüber hinaus konstant erweitert. Bezüglich Arbeitspaket 4 «Automatische Übersetzung» zeigte sich, dass die lokal in Chur verfügbaren Rechenressourcen ausreichen um ein Übersetzungsmodell mit der verfügbaren Datenmenge zu erstellen. Auf Meilenstein 4 «Erstes Training mit CSCS» konnte deshalb verzichtet werden. Konstante Evaluierungen im Arbeitspaket 5 «Prototyp und Evaluation» stellten die Qualitätsentwicklung sicher. Zur Unterstützung wurde hierbei eine weitere Mitarbeiterin mit muttersprachlichen Kenntnissen der Rätoromanischen Sprache hinzugezogen. Das Arbeitspaket 6 «Dissemination» konnte mit der Aufschaltung eines Web-Portals, welches freien Zugang zum Übersetzungsdienst gewährt, abgeschlossen werden. Über den Projektrahmen hinaus sind weitere Tätigkeiten zur Bekanntmachung geplant.

3) Resultate/ Ergebnisse:

Es konnte aufgezeigt werden, dass die Entwicklung von Technologien zur Unterstützung der Rätoromanischen Sprache durchaus möglich ist. Konkret wurde ein Web-Portal geschaffen, welches es erlaubt computerbasiert beliebigen Text von Deutsch ins Rätoromanische zu übersetzen. Wie erwartet ist die Qualität dieser Übersetzungen noch unzureichend, um eine Effizienzsteigerung in darauf aufbauenden Prozessen zu erreichen, allerdings zeigt sich sehr wohl das Potential, welches der gewählte Ansatz verspricht. So liegt die Leistungsfähigkeit des Übersetzungsdienstes bei 39 BLEU-Punkten – etwa halb so gut wie ein menschliches Pendant und auf gleichem Level wie vergleichbare Forschungsergebnisse in anderen Sprachen.

Im Projektverlauf wurden verschiedene Trainingsdatenkorpora gebildet, welche sich in ihrer Nutzbarkeit unterscheiden. So konnten wir 141'836 Übersetzungen aus öffentlich zugänglichen Quellen (Gesetzestexte, Medienmitteilungen, Convivenza-Blog) sammeln und geeignet aufbereiten, welche anderen Forschungsgruppen ohne Bedingung weitergegeben werden können. Ein weiterer Korpus beinhaltet Übersetzungen und Textdokumente aus Firmendaten, welche ausschliesslich im Rahmen der Arbeit der Forschungsgruppe verwendet werden darf und somit als Alleinstellungsmerkmal der entwickelten Forschungsk Kooperation dient. Ein dritter Korpus soll Unternehmensdaten enthalten, welche in anonymisierter Form als Korpus anderen Forschungsgruppen zur Verfügung gestellt werden kann.

Ein weiteres wichtiges Ergebnis ist die im Projekt etablierte Forschungsk Kooperation Bündner Institutionen zur Förderung der Rätoromanischen Sprache mittels fortgeschrittener KI-Technologien. Ebenfalls konnte über das Projekt hinaus die Einführung einer Computer Assisted Translation Software beim Übersetzungsdienst der Lia Rumantscha angestossen werden. Dies erlaubt die effizientere Übersetzung und Generierung von Übersetzungsdateien in einem standardisierten Format, sodass in Zukunft keine zusätzliche Aufbereitung entsprechender Daten mehr erforderlich ist.

4) Schlussfolgerungen/ Synthese:

Es wurde gezeigt, dass die FHGR über die benötigten Kompetenzen zur Entwicklung von Übersetzungstechnologie für die Rätoromanische Sprache verfügt. Das Ergebnis des Forschungsprojekts eignet sich zwar eingeschränkt für einfache Übersetzungen in bestimmten Domänen, aber – wie erwartet – selbst noch nicht dafür, Unternehmensprozesse effektiv zu optimieren. Allerdings konnte die generelle Umsetzbarkeit gezeigt und eindrucksvoll sowohl das Potential als auch die Herausforderungen unterstrichen werden, welche die weitere Entwicklung dieser Technologie begleiten. Die Erkenntnisse und der Prototyp werden dafür genutzt um für eine Folgefinanzierung der Forschung in dieser Disziplin zu werben. So wird beabsichtigt, eine entsprechende Eingabe beim Forschungsfond GRDigital zu tätigen.

Die gewonnenen Erkenntnisse erlauben eine umfassende und exakte Einschätzung von Aufwand, Herausforderungen und Synergiepotentialen und unterstützen somit sofort in der Planung von darauf aufbauenden Arbeiten.

Die Verwertung der Ergebnisse in Peer Reviewed Publikationen wird derzeit vorbereitet. Der Kompetenzaufbau kommt unter anderem dem neuen Studiengang "Computational and Data Science» (CDS) zugute, in dem Deep Translation als Vertiefungsgebiet und Anwendungsfach gelehrt wird.

5) Kosten/ Finanzen Überblick über das Gesamtprojekt:

a) Einnahmen:

Gesprochener Beitrag Förderverein	CHF 40'000
Zuwendung in Form von Arbeitsleistung (Lia Rumantscha)	CHF 30'000
Total	CHF 70'000

b) Ausgaben:

Personalkosten FHGR	CHF 40'514.5
Personalkosten Lia Rumantscha	CHF 30'000
Total	CHF 70'514.5

c) Kostenstand:

Das Projekt wurde mit einem Defizit von CHF 514.50 abgeschlossen. Die Auszahlung des bewilligten Betrags von CHF 40'000 durch den Förderverein steht aus.

6) Publikationen:

Es besteht noch keine nachweisbare Publikationsleistung. Zu erwarten sind einzelne Bekanntmachungen in den lokalen Medien, sowie Peer Reviewed Konferenzbeiträge an Fachtagungen.