Verbeurde geleenthede in vertaalgeheuepassing

Author:

Friedel Wolff¹

Affiliation:

¹School for Computer Science, University of South Africa, South Africa

Correspondence to:

Friedel Wolff

Email:

wolfff@unisa.ac.za

Postal address:

PO Box 386, University of South Africa 0003, South Africa

How to cite this abstract:

Wolff, F., 2015, 'Verbeurde geleenthede in vertaalgeheuepassing', *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 34(1), Art. #1333, 1 page. <http://dx.doi.org/10.4102/satnt.v34i1.1333>

Note:

A selection of conference proceedings: Student Symposium in Science, 06 and 07 November 2014, Science Campus, University of South Africa. Organising committee: Mr Rudi W. Pretorius and Ms Andrea Lombard (Department of Geography, University of South Africa) and Dr Hertzog Bisset (South African Nuclear Energy Corporation [NECSA]).

Copyright:

© 2015. The Authors. Licensee: AOSIS OpenJournals. This work is licensed under the Creative Commons Attribution License.

Missed opportunities in translation memory matching. In an attempt to improve translation memory systems, this study investigates which useful suggestions are not suggested by current methods. It classifies results from two language pairs in two data sets, and identifies shortcomings relating to semantics, omission, text normalisation and misalignment.

'n Vertaalgeheue is 'n databasis waarin 'n gebruiker op 'n gestruktureerde manier voorheen vertaalde (teken-)teks kan stoor met ooreenstemmende bronteks. Vertaalgeheues word algemeen gebruik en verbeter vertalers se produktiwiteit en konsekwentheid. 'n Vertaalgeheuestelsel funksioneer as 'n inligtingherwinningstelsel wat een of meer voorstelle uit die vertaalgeheue probeer onttrek wat die vertaler in sy huidige vertaaltaak sal bystaan. 'n Verbetering in die herroepingsvermoë van só 'n stelsel is wenslik. Hierdie studie fokus op daardie voorstelle wat nuttig sou wees, maar wat nie deur 'n basislynmate vir soortgelykheid gekies word nie.

Gedurende normale werking soek 'n vertaalgeheuestelsel deur die gestoorde vertaalpare vir brontekste soortgelyk aan die huidige segment wat vertaal word. Die mate vir soortgelykheid is tipies iets soos die Levenshtein-afstand. Indien sulke brontekste gevind word, word die ooreenstemmende teikenteks vir die vertaler aangebied om te help met die vertaling van die huidige segment. 'n Nuttige voorstel hoef slegs effens geredigeer te word (of glad nie) om dit te omskep in die verlangde vertaling in die teikentaal. Ander teikensegmente bestaan moontlik in die vertaalgeheue wat ewe nuttig kan wees, maar wat nie gekies word nie omdat hulle bronteks nie soortgelyk genoeg is aan die navraag (huidige segment wat vertaal word) nie.

In 'n poging om vertaalgeheuestelsels te verbeter, probeer hierdie studie om dié vraag te beantwoord: Wat veroorsaak dat nuttige idees nie voorgestel word nie, en hoe groot is die verbetering wat steeds moontlik is?

In 'n neutredop: oorweeg alle segmente in 'n gegewe vertaalgeheue en identifiseer hoogs soortgelyke teikensegmente, oorweeg dan hul ooreenstemmende bronsegmente en kies daardie pare waarvoor die soortgelykheid van die bronsegmente onder die nodige drempel is. Dié pare verteenwoordig die 'verbeurde geleenthede'. Dié verbeurde geleenthede word dan handmatig ondersoek om te verstaan hoekom hulle nie deur die basislynbenadering gekies word nie. In elke geval word die twee brontekste vergelyk om 'n ooglopende verwantskap tussen hulle te identifiseer wat kan verklaar hoekom die teikentekste soortgelyk is. Deur hierdie verwantskappe te kategoriseer en te kwantifiseer, kry mens 'n aanduiding van waar fokus benodig word om die herroeping van vertaalgeheuestelsels te verbeter.

Twee taalkundig onverwante taalpare uit twee heeltemal verskillende datastelle word ondersoek.

Vier hoofkategorieë word geïdentifiseer waarvolgens verbeurde geleenthede verder ontgin kan word. Die grootste kategorie het te doen met semantiek (gevalle met sinonieme, parafrases, afkortings, ensovoorts). In ander gevalle is weglating, teksnormalisering en swak belyning waargeneem.

Hierdie studie bespreek die geïdentifiseerde kategorieë en moontlike redes waarom hulle nie gekies word nie. Daar is ook 'n bespreking van 'n paar moontlike oplossings vir die verbetering van die huidige stelsels.

Read online:

Scan this QR code with your smart phone or mobile device to read online.