Grover, A.S.; van Huyssteen, G.B.; Pretorius, M.W.;
Human Language Technol. Res. Group, Meraka Inst., South Africa

**This paper appears in:** Technology Management in the Energy Smart World (PICMET), 2011 Proceedings of PICMET '11:
**Issue Date:** July 31 2011-Aug. 4 2011
**On page(s):** 1 - 14
**Location:** Portland, OR
**E-ISBN:** 978-1-890843-24-3
**Print ISBN:** 978-1-4577-1552-5
**INSPEC Accession Number:** 12229400
**Date of Current Version:** 12 September 2011

## ABSTRACT

South Africa (SA) epitomises diversity, with the nation boasting eleven official languages. The field of human language technology (HLT) can play a vital role in bridging the digital divide and thus has been recognised as a priority area by the South African government. The current HLT landscape in South Africa consists mostly of a relatively young research and development (R&D) community, the government and a handful of private sector companies. A key challenge is the perceived fragmentation of the R&D activities in this domain; there is insufficient codified knowledge about the currently available South African HLT language resources (LRs) and applications. In this paper we describe a national technology audit we undertook for the South African HLT landscape. The objective of our study was to codify and present a profile of HLT components in the South African HLT R&D environment. We present the technology audit process employed, which involved various data collection methods such as expert consultations, workshops and questionnaires. We also describe the complementary approaches used to analyse the status of the landscape, such as the detailed inventories of HLTs available across South Africa's eleven languages and a series of indexes developed to provide a landscape overview. We found that a number of HLT LRs are available in South Africa but are of a very basic and exploratory nature and there are many areas that lie fallow in terms of the variety, number, technology maturity and accessibility of HLT items.