

io-port 05677921

Lindén, Krister; Silfverberg, Miikka; Pirinen, Tommi

HFST tools for morphology — an efficient open-source package for construction of morphological analyzers.

Mahlow, Cerstin (ed.) et al., State of the art in computational morphology. Workshop on systems and frameworks for computational morphology, SFCM 2009, Zurich, Switzerland, September 4, 2009. Proceedings. Berlin: Springer (ISBN 978-3-642-04130-3/pbk; 978-3-642-04131-0/ebook). Communications in Computer and Information Science 41, 28-47 (2009).

Summary: Morphological analysis of a wide range of languages can be implemented efficiently using finite-state transducer technologies. Over the last 30 years, a number of attempts have been made to create tools for computational morphologies. The two main competing approaches have been parallel vs. cascaded rule application. The parallel rule application was originally introduced by Koskenniemi and implemented in tools like TWOLC and LEXC. Currently many applications of morphologies could use dictionaries encoding the a priori likelihoods of words and expressions as well as the likelihood of relations to other representations or languages. We have made the choice to create open-source tools and language descriptions in order to let as many as possible participate in the effort. The current article presents some of the main tools that we have created such as HFST-LEXC, HFST-TWOLC and HFST-Compose-Intersect. We evaluate their efficiency in comparison to some similar tools and libraries. In particular, we evaluate them using several full-fledged morphological descriptions. Our tools compare well with similar open source tools, even if we still have some challenges ahead before we can catch up with the commercial tools. We demonstrate that for various reasons a parallel rule approach still seems to be more efficient than a cascaded rule approach when developing finite-state morphologies.

doi:10.1007/978-3-642-04131-0_3