

Unitag for Nepali

The enclosed programs form an instantiation of the Unitag system (see Hardie 2004) designed for POS-tagging of Nepali text.

The binaries supplied are Windows-32 binaries only. For source code that can be compiled on other systems, please see <http://unitag.sourceforge.net/>. Unitag was created by Andrew Hardie (<http://www.ling.lancs.ac.uk/staff/hardie>).

Input format

The system requires 16-bit Unicode using the native byte ordering for the operating system (this is LE for Windows systems; on other systems it may be BE). Both plain text and XML can be processed.

System calls

To run Unitag for Nepali, enter

```
unitag nepali-inst.txt raw_text_filename
```

where the second argument is the filename of the input file; or

```
unitag nepali-inst.txt listfile_filename L
```

where the second filename is an ASCII file containing a list of input files, one file per line. The separate *deverticalise* program converts the output to XML: the call for *deverticalise* is:

```
deverticalise unitag_tagged_file output_xml_file
```

Licensing and source code availability

Unitag is free software and its source code is available at <http://unitag.sourceforge.net/>

Unitag is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 3 of the License, or (at your option) any later version.

Unitag is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Reference

Hardie, A (2004) The computational analysis of morphosyntactic categories in Urdu. PhD thesis, University of Lancaster. <http://eprints.lancs.ac.uk/106/>