IRSTLM Toolkit by FBK

The META Technology Council proudly award the META Seal of Recognition to

IRSTLM Toolkit by FBK

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Bernardo Magnini
(on behalf of Marcello Federico)

www.meta-net.eu
A toolkit for the estimation, representation, and computation of statistical language models

• advanced smoothing methods for large LMs
• approximate smoothing for gigantic LMs
• pruning and quantization of LMs
• efficiently storing LMs on disk and in memory
• several LM adaptation methods
• compatible with LMs created with other tools
• integrated inside MOSES and FBK’s ASR
• open source LGPL license:

http://sourceforge.net/projects/irstlm
Evolution

- 1992: Command and control of a robot (1K words)
- 1995: Dictation of medical reports (10Kw)
- 1997: Spoken data entry (10Kw, recursive LMs)
- 2000: Broadcast news transcription (60Kw, 5M 3-grams)
- 2004: News translation (2Mw, 200M 4-grams)
- 2006: Translation of speeches (4.5Mw, 350M 5-grams)
- 2008: NIST MT task (6G 6-grams, on-disk creation)
- 2011: Translation of TED talks (multiple, hybrid LMs)
• more than 10,000 downloads
• many from companies
  Apptek, Applied Language Solutions, Translated, Pervoice,
  Digital Silk Road, Symantec Ltd, Asia Online, …
• and public institutions
  JRC and DGT (EC), World Intellectual Property Organization (UN)
Credits

Marcello Federico
Nicola Bertoldi
Mauro Cettolo

Chris Dyer
Fabio Brugnara
Arianna Bisazza
Christian Hardmeier
Hieu Hoang

IRSTLM