XFST by Xerox

THE META TECHNOLOGY COUNCIL PROUDLY AWARDS THE META SEAL OF RECOGNITION TO

XFST by Xerox

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

Lauri Karttunen (video)

On behalf of the META Technology Council

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

www.meta-net.eu

On behalf of the META Technology Council

Georg Rehm
Network Manager of META-NET
DFKI, Germany

XFST
Xerox Finite-State Toolkit

Lauri Karttunen

with

Ronald M. Kaplan, Todd Yampol, Pasi Tapanainen,
Kenneth R. Beesley, Hervé Poirier, André Kempe,
Tamás Gaál, and many others

1993 – present and future

parc
Palo Alto Research Center

What is XFST?

- A development tool for compiling finite-state networks from
  - strings
  - text files
  - xfst regular expressions

  with calculus operations (union, intersection, composition,…)

- A utility for applying finite state networks to text for
  - rewriting
  - tokenization
  - morphological analysis and generation
  - pattern matching

parc
Palo Alto Research Center

What is XFST used for?

- Since its inception in 1993 XFST has been used to create lingware for dozens of languages
  - English, Finnish, French, German, Italian, Spanish, …
  - Arabic, Basque, Georgian, Irish, Turkish, …
  - Mongolian, Rotokas, Sámi, Zulu, Xhosa, …

- Industrial users of XFST and its commercial version include
  - Temis
  - Basis Technology
  - SAP
  - Microsoft

parc
Palo Alto Research Center

Future of XFST

- In the near future Xerox will release the source code of XFST and its enhanced industrial version freely for non-commercial use under a click-through license.

- The current executable is available at

  http://www.stanford.edu/~laurik/.book2software/

- Other resources:

  Finite State Morphology

  Kenneth R. Beesley and Lauri Karttunen

  CSLI Publications, 2003

  http://www.fsmbook.com/

parc
Palo Alto Research Center