



# Sharing Language Technology Information

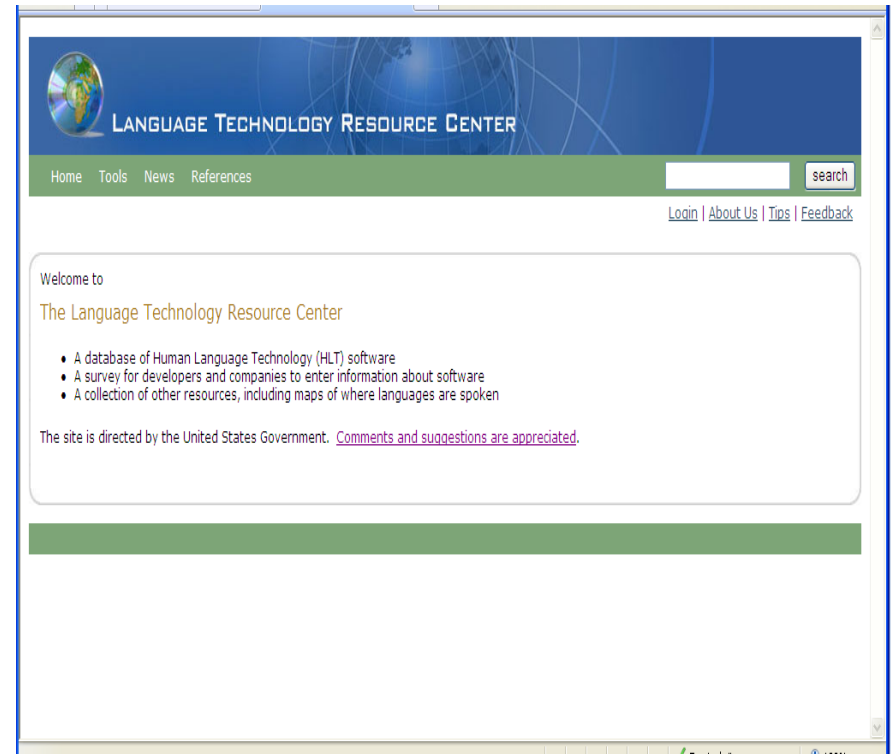
**EAMT  
May 2010**

**Dr. Jennifer DeCamp  
MITRE**

**[jdecamp@mitre.org](mailto:jdecamp@mitre.org)**

# Language Technology Resource Center

- Web site regarding language technology
  - MITRE
  - U.S. Government
- Started in 2000
- Survey and reports
- User information
  - FAQs
  - Guides
- Links to online information
- Standards information
- Meetings and training
- SharePoint, wikis, blogs
- U.S. government outreach
- Just LTs—not LRs



<https://ltrc.mitre.org>

# Product Comparisons

- Meets requirement
- Does not meet requirement

Product Name	Meets all hard requirements	Process in batch mode	Enhance images or deal with poor quality images	Recognize charts	Replace digital text in charts	Recognize tables	Replace digital text in tables	Recognize columns	Recognize handwriting	Learn	Provide high accuracy	Cost
<a href="#">Sakhr OCR</a>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	TBD	1400
<a href="#">ArborScript CRS</a>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	TBD	Quotes available upon request
<a href="#">VERUS</a>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	TBD	VERUS Standard: \$1299
<a href="#">Automatic Reader Arabic OCR</a>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	TBD	1499
<a href="#">BBN Optical Character Recognition Technology</a>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	TBD	

H ▾ H ▾ H ▾ H ▾ H ▾ H ▾ H ▾ H ▾ H ▾

Recalculate Scores

# Collaborative Space

The screenshot displays a SharePoint site interface for 'Dari Working Group'. The top navigation bar includes 'Home', 'Find Tools', 'Forums', and 'Events'. A search bar is present with the text 'This Site: Dari Working Group'. The main content area is divided into several sections:

- Announcements:** A message stating 'There are currently no active announcements. To add a new announcement, click "Add new announcement" below.'
- Discussion Forum Posts and Responses:** A section with a table header: 'Subject', 'Created By', 'Replies', and 'Last Updated'. Below the header, it says 'There are no items to show in this view of the "Discussion Forum Posts and Responses" discussion board. To create a new item, click "New" above.' and includes an 'Add new discussion' button.
- Document Library:** A section with a table header: 'Type', 'Name', and 'Modified By'. Below the header, it says 'There are no items to show in this view of the "Document Library" document library. To create a new item, click "Add new document" below.' and includes an 'Add new document' button.
- Dari Working Group Resources:** A section with the text 'There are currently no favorite links to display. To add a new link, click "Add new link" below.' and an 'Add new link' button.
- Calendar:** A section with the text 'There are currently no upcoming events. To add a new event, click "Add new event" below.' and an 'Add new event' button.

A left-hand navigation pane contains the following categories:

- View All Site Content
- Documents
  - Document Library
- Lists
  - Calendar
  - Tasks
- Discussions
  - Discussion Forum Posts and Responses
- Sites
- People and Groups
  - Recycle Bin

# Information on Working With the U.S. Government

- **Finding out about U.S. Government work**
  - **Informing the U.S. Government about your work**
  - **Finding out about grants**
-

# Issues

- **Terminology**
  - **Separation and aggregation**
  - **Different requirement sets**
  - **Comparisons**
  - **Need for user education (translators, language teachers, etc.)**
  - **Sensitivity of research and industry information**
  - **Validation and testing**
  - **Incomplete or outdated information**
  - **Leveraging and collaboration**
-

# Current Work

- **Soliciting input**
  - **Updating survey**
    - Including more on capabilities
    - Including ability for multiple sources to add information, with sources marked
  - **Developing guides**
  - **Connecting with other efforts**
-

# Related Efforts

- **ACL's Natural Language Software Registry**
  - **CLARIN**
  - **Compendium for Translation Software**
  - **FLaReNet**
  - **MetaNet**
  - **Others**
-



# Questions

- **How can we share information to avoid duplicate work?**
    - **Developers filling out surveys**
    - **Staff writing summaries**
  - **What set of information should we look at?**
  - **How detailed should this information be?**
  - **To what extent can we automate it?**
  - **How do we validate information?**
  - **How can we provide data-specific testing?**
  - **How do we keep information updated?**
  - **How do we communicate the range of what we cover?**
    - **Citing sources**
    - **Describing additional sources**
-