Sharing Language Technology Information

EAMT
May 2010

Dr. Jennifer DeCamp
MITRE
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

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Meets all hard requirements</th>
<th>Process in batch mode</th>
<th>Enhance images or deal with poor quality images</th>
<th>Recognize charts</th>
<th>Replace digital text in charts</th>
<th>Recognize tables</th>
<th>Replace digital text in tables</th>
<th>Recognize columns</th>
<th>Recognize handwriting</th>
<th>Learn</th>
<th>Provide high accuracy</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sakhr OCR</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>TBD</td>
<td>1,400</td>
</tr>
<tr>
<td>ArborScript CRS</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
<td>○</td>
<td>TBD</td>
<td>Quotes available upon request</td>
</tr>
<tr>
<td>VERUS</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>TBD</td>
<td>VERUS Standard: $1,299</td>
</tr>
<tr>
<td>Automatic Reader Arabic OCR</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>TBD</td>
<td>1,499</td>
</tr>
<tr>
<td>BBN Optical Character Recognition Technology</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>TBD</td>
<td></td>
</tr>
</tbody>
</table>

Recalculate Scores
Collaborative Space
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- Incomplete or outdated information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