A Cloud-based Strategy for Facilitating Adoption of Open-Source NLP in Applied Research Settings

David Carrell

Cloud Deployment Model

Developer Institution 1

Developer Institution N

Cloud Computing solves several technical challenges and introduces a major new challenge: achieving security outside the local institutional firewall. Addressing this will require:

- Local stakeholder participation in security assessment, planning, implementation, auditing, etc.
- Low-risk pilot opportunities to prove the concept and build confidence in security measures (e.g., processing 100% de-identified text, initially).
- Building a constituency of local researchers who recognize the potential advantages of local NLP capacity.
- Sacrificing NLP system performance in the interest of reducing risk exposure (e.g., not persisting clinical text in the cloud).
- Education of stakeholders.

Developer Institutions

In this collaborative model one or more developer institutions assumes responsibility for:

- Deploying to the cloud the NLP system and a locally-deployed “I/O Manager” application that interfaces with the local text repository at the user site.
- Updating/tweaking the NLP pipeline as needed to address specific information extraction tasks, bugs.
- Advising the user sites on pre-processing, post-processing, and validation tasks performed locally at each user site.

Developer responsibilities may be coordinated and shared among multiple institutions.