Welcome to Scrubber

Patient Text De-identification, Natural Language Processing

See **Apache cTAKES** for the last developer resources.

Scrubber helps investigators ensure patient privacy when using physician notes for clinical research. Scrubber removes HIPAA defined Protected Health Identifiers by matching human expert and and machine defined rules for text processing. This software can work with note text either unformatted or contained within XML files or SQL databases. Launched in 2006, the use of Scrubber has been approved by numerous hospital IRBs and quality has been validated by physician review.

Update OCT 16, 2013

New Scrubber paper with Apache cTAKES

> ⓘ "Improved de-identification of physician notes through integrative modeling of both identifying and non-identifying medical text", McMurry* AJ, Fitch* B, Savova G, Kohane IS, Reis BY.BMC Medical Informatics and Decision Making

Update JUNE 18, 2013

Scrubber pipeline is moving to Apache cTAKES! .
For recent progress, see this presentation to the I2b2 National Center for Biomedical Computing:

> ⓘ "Integrating Public and Private Medical Texts for Patient De-Identification with Apache cTAKES"