PDF

XML

OCR

Document Structure

SciXML

Named Entity Recognition (genes)

Sentence Parsing (RASP)

Anaphora Resolution

FBXML (xml containing our own “added value” markup)

PaperBrowser & navigators