embed\_0

Apache Tika - a content analysis toolkit



detects and extracts metadata and text