Self-service plagiarism checking tool for non-profit research institute

CHALLENGES

The research institute needed an at-scale plagiarism checker to better review articles before publication. They needed a solution that could be a self-service, browser-based, easy to use internal tool.

SOLUTIONS

Neal Analytics built an Azure Data Lake host solution and secure access with Active Directory Federation Services (ADFS). We used Azure Cognitive Services for Optical Character Recognition (OCR) on non-text files like PDF, JPG, TIFF, and the Bing API for plagiarism checking.

SOLUTIONS

The results of this sentence-level search over the whole document content are aggregated and an aggregate plagiarism index is created per URL. For instance, the tool would indicate that 24% of the document is a plagiarism of URL A, and 5% from URL B.

In addition, Neal also developed an easy-to-use, web-based user interface. Users navigate to this internal webpage, upload their document, and get plagiarism results with a simple click.

RESULTS

Using the Azure Data Lake host solution, any authenticated user can upload documents in pdf, image or txt format, from any browser. To help with plagiarism checking sensitivity, the tool also has a confidence level parameter that is, by default, set at 60%.

Azure Cognitive Services and Web-based UX provided:

- % text likely coming from existing sources (plagiarism)
- Top 10 sources (with associated %) URLs
- Color-coded sentences associated with URLs
- Export as pdf of the full report, including color-coding and URLs.

11911 NE 1st Street, Suite 206, Bellevue, WA 98005 | 425.822.2332 | contactus@nealanalytics.com

@NealAnalytics LinkedIn.com/company/NealAnalytics