Software as a Service - Solutions for Word Processing

John Wensink

MIS440 Cloud Computing and Big Data

Colorado State University-Global Campus

Dr. Mazen Alkhatib

January 26, 2020

Software as a Service - Solutions for Word Processing

In today’s robust world of cloud computing, Software as a Service (SaaS) describes a distribution model in which a remote host provides an application that can be used anywhere that there is web connectivity. SaaS differs from on-premises computing, in that nearly all resources are maintained by the vendor of the application. All the user has to do is connect to the service without regard to maintaining the servers and other computational resources required to provide such a service in order to use that resource’s functionality. In this paper, we will look at two prolific SaaS applications related to the creation of text documents.

**Google Docs**

Google Docs is part of the Google Drive suite of office productivity tools. The origin of this web application came from two startup companies, both of which were acquired by Google in 2009 and 2010. Writely is a browser-based document creation tool that was released in August of 2005 by the Upstartle company. Writely was unique for its time in that it featured an Ajax-based word processor that allowed a browser application to function similarly to a desktop-based application. Ajax stands for Asynchronous JavaScript XML, a client-side development technique that allowed a document to be processed and manipulated before its transmission to the host server (W3Schools, n.d.). Writely was not a true cloud-based application, as one would be required to download the application to their computer before use. The other predecessor to Google Docs was named XL2Web, a dynamic application that allowed content (such as complex documents and spreadsheets) to be converted into XML-based files. This conversion would then would allow for the possibility for those files to be published to, and retrieved from JEngine, a type of server specifically created to host Java applications and the content they create. Google acquired the company 2Web Technologies including XL2Web in 2005 (Rochelle, 2019), and Upstartle with its product Writely in March of 2006 (Marshall, 2006). By July of 2009, Google Docs was ready to conclude beta testing and proceed to mass deployment of its product (Google, 2009). The web-based document creator was meant to compete with Microsoft Word, Apache OpenOffice, and LibreOffice, and includes features that one would expect from an application designed to create textual documents. The difference is that Google Docs is entirely hosted in the cloud. No software is required to be downloaded, and documents created are saved remotely on Google’s servers (although local copies can be downloaded.) Google Docs is not limited to simply creating documents, the service can be used to store a vast collection of third party documents and with the combination of Document AI, business intelligence can be gathered from the data mine.

**Unstructured Data and Machine Learning**

Google Docs is able to be used to analyze unstructured data such as pictures, audio, and video through the use of the open-source TensorFlow machine learning toolkit. TensorFlow was originally created for internal use only by Google but now is available to use for big-data analytics to the general population (Metz, 2018). The second iteration of Google’s DistBelief, TensorFlow utilizes machine learning based on deep neural nets that can be used on a single computer but yet is scalable and able to be used across multiple processors. In computer science, a tensor is jargon for an operation that is performed on multidimensional (more than two) data arrays (MathWorks, n.d.). Google Docs uses Tensorflow to classify unstructured data using the process of learning by example. Using artificial neural networks that were inspired by the structure of biological neural systems, TensorFlow uses techniques such as activation functions (non-linear transformations) on an input signal such as an image or other non-relational data. Using the example of image classification, Google’s Document AI is able to draw meaningful insights regarding unstructured data in large document sets by classifying images, turning the unstructured data into structured data, so that insight may be created where previously impossible (Google, n.d.). Another example of how Google’s Document AI uses TensorFlow to improve business intelligence using unstructured data is that documents in over 200 different languages can be recognized using natural language processing techniques in order to classify previously unrelated datasets (Google, n.d.). The software is so impressive that it can provide these insights even when processing handwritten images of words in different languages. Remarkably, in November of 2015, Google released TensorFlow under the open-source Apache 2.0 license (Metz, 2018).

In May of 2017 Google Docs was the target of a spearphishing campaign when a series of emails camouflaged as a genuine Google Docs invitation (from a known contact) requested document sharing permissions from the targeted individual (Levin, 2017). When the malicious link was clicked, a request was issued to grant permission to the target’s Google account. The request appeared to be from Google Docs, however, in actuality, it was from the actor attempting to gain access to the target’s Gmail account. This was possible because the request came from a page stored on Google’s server domain. Once permission was granted, the malware used email contact information to send out more targeted emails to new victims. On the very same day that the attacks were detected, Google took action to prevent further damage by removing the attacker’s page in which the requests came from, pushed an update for Android’s version of Gmail, and sent notifications to potential victims alerting them to the possibility of this attack.

**Grammarly**

The second SaaS product that will be discussed is a Ukranian writing tool called Grammarly. Grammarly also uses deep natural language processing algorithms, combined with machine learning to create a web-based tool that helps document creators improve their quality of writing. Grammarly was launched in July of 2009 from its headquarters in San Francisco California. The company also maintains offices in New York City, and Vancouver, and Kiev Ukraine. If you have ever used a spelling checker when creating a word document, you will have an idea of how Grammarly’s graphical user interface works. A spelling/grammar error, or an area where one’s writing could be improved is highlighted and underlined to flag a potential problem with the text. When a writer floats their cursor over the highlighted text, suggestions are given to correct spelling, grammar, and poor/overused sentence structure. “Grammarly’s algorithms flag potential issues in the text and suggest context-specific corrections for grammar, spelling and usage, wordiness, style, punctuation, and even plagiarism. Grammarly explains the reasoning behind each correction, so you can make an informed decision about whether, and how, to correct an issue” (Grammarly, n.d. a).

**Social/Collaborative API and Data Security**

Social features inside Grammarly include collaborative file creation, revision history, and actionable items for individual authors based on suggestions from other authors collaborating on the same document. Visualization tools inside Grammarly allow groups to see which individual was responsible for creating which blocks of text, by presenting them in different colors unique to the author who was responsible. Grammarly’s API has been included as a default extension on Google Docs when using the Chrome web browser, it can be found on Grammarly’s website as a web-based SaaS, as well as a standalone program available to download on individual systems.

Grammarly relies on Amazon Web Services (AWS) for its customers in the United States, and all data is processed in a private cloud operated by Grammarly. The service uses segregated servers, firewalls, and load balancers to ensure that data sent to Grammarly remains secure. Grammarly uses AES-256 encryption to store data at rest, and TLS 1.2 to encrypt data in motion. Grammarly employees undergo mandatory training in areas such as “data privacy, physical security, data and information security, and incident reporting” (Grammarly, n.d. b). Grammarly also offers a bug bounty program where ethical hackers can present issues to the company for a cash reward. In 2018, a security vulnerability was discovered in Grammarly’s Chrome extension that allowed attackers to hijack a user’s session and display all text that had ever been typed into the browser when using the Grammarly editor. Although Grammarly contends that this exploit was rapidly patched and no user data was compromised, the severity of this type of exploit has prompted Grammarly to take digital security more seriously (O’Niell, 2019).

Software as a Service has grown at an incredible pace in the last ten years. SaaS has already become a staple of the modern internet. Combined with advances in artificial intelligence, machine learning, and deep neural networks, the next ten years are sure to have some amazing capabilities yet to be discovered.

References

Google. (n.d.). Document AI | Google Cloud. Retrieved January 26, 2020, from <https://cloud.google.com/solutions/document-ai>

Google. (2009, July 7). Google Apps is out of beta (yes, really). Retrieved January 26, 2020, from <https://googleblog.blogspot.com/2009/07/google-apps-is-out-of-beta-yes-really.html>

Grammarly. (n.d. a). How does Grammarly work? Retrieved January 26, 2020, from <https://support.grammarly.com/hc/en-us/articles/115000090871-How-does-Grammarly-work->

Grammarly. (n.d. b). Security at Grammarly. Retrieved January 26, 2020, from <https://www.grammarly.com/security-practices>

Levin, S. (2017, May 3). Google Docs users hit with sophisticated phishing attack in their inboxes. Retrieved January 26, 2020, from <https://www.theguardian.com/technology/2017/may/03/google-docs-phishing-attack-malware>

Marshall, M. (2006, March 10). Google acquires online word processor, Writely. Retrieved January 26, 2020, from <https://venturebeat.com/2006/03/09/google-acquires-online-word-processor-writely/>

Metz, C. (2018, March 6). Google Just Open Sourced the Artificial Intelligence Engine at the Heart of Its Online Empire. Retrieved January 26, 2020, from <https://www.wired.com/2015/11/google-open-sources-its-artificial-intelligence-engine/>

O'Neill, P. H. (2019, February 28). Bug in Grammarly browser extension exposes what a user ever writes. Retrieved January 26, 2020, from <https://www.cyberscoop.com/bug-in-grammarly-browser-extension-exposes-virtually-everything-a-user-ever-writes/>

MathWorks. (n.d.). Multidimensional Arrays. Retrieved January 26, 2020, from <https://www.mathworks.com/help/matlab/math/multidimensional-arrays.html>

Rochelle, J. (2019). Jonathan Rochelle. Retrieved January 26, 2020, from <https://www.jonathanrochelle.com/>

W3Schools. (n.d.). What Is AJAX? Retrieved January 26, 2020, from <https://www.w3schools.com/whatis/whatis_ajax.asp>