

# Appraise on Azure: A cloud-based, multi-purpose evaluation framework

Christian Federmann  
Microsoft Research  
One Microsoft Way  
Redmond, WA 98052, USA  
chrife@microsoft.com

## Abstract

We present Appraise on Azure, an extension to the Appraise evaluation framework which enables users to host evaluation campaigns in the Microsoft Azure cloud. This allows to scale annotation efforts on demand and makes it easier to set up and run multiple annotation tasks in parallel. Both the Appraise framework and the code adding Azure support are released under an open license. The demo will give details on the architecture of the system, discuss Azure integration and demonstrate annotation options in the framework.

## 1 Motivation

Over the last years, the Appraise evaluation toolkit has seen growing interest from industry and research community. This has resulted in a number of forks of the source code which is publicly available.<sup>1</sup> Work on the updated version presented in this demo was motivated by problems with the original version:

- **Outdated forks:** As the underlying software packages have been updated over time, the integration of external code is difficult. Potential security issues might remain unfixed.
- **No support for parallel campaigns:** The original version of Appraise was implemented as an open framework which allowed to build annotation platforms for specific scenarios. However, the framework lacked support to run multiple such collection efforts in parallel.

- **Too WMT centric:** Appraise is the platform for the yearly evaluation campaign conducted as part of the WMT Conference on Statistical Machine Translation. Hence, focus shifted from supporting a wide range of annotation tasks to only two: relative ranking and, later, direct assessment. Also, the software became harder to configure for non-WMT campaigns, adding unnecessary complexity to the setup.

## 2 Appraise on Azure Improvements

We have addressed the aforementioned issues with the old codebase and added support for the latest Django version 1.11. Also, we implemented support to host Appraise on Microsoft Azure. The Microsoft Azure cloud<sup>2</sup> is a collection of integrated cloud services, including hosting, storage and compute solutions as well as high-level APIs such as Cognitive Services. The simplified setup process makes it easier to focus on the creation of new annotation views. The same holds if a user only wants to run an evaluation campaign. The end user benefits in multiple ways: 1) all software dependencies have been updated. 2) it is now possible to configure multiple annotation campaigns in a single Appraise instance. 3) users can set up non-WMT tasks while still getting updated status views and other features introduced for WMT.

## 3 Conclusion

Our demo presents Appraise on Azure, a cloud-based, multi-purpose evaluation framework. We discuss the architecture, annotation views and give an outlook to future extensions. All features work both standalone and when hosting Appraise on Microsoft Azure.

<sup>1</sup> See <https://github.com/cfedermann/Appraise/>

<sup>2</sup> See <https://azure.microsoft.com/>