Recommendation Engine

<https://kabroo.visualstudio.com/DTML.org/_workitems/edit/1131>

Status: Draft

Introduction

We are looking to improve the current recommendation engine used by DTML to more effectively assign users words to learn. Our goal is to make the recommendation engine more effective at growing the user’s language proficiency, and to reduce the load on DTML servers by using user data to pre-calculate a set of words to be recommended on a per-user basis.

Short description

We will be developing a modified representation of storing word similarity and difficulty. Additionally, we will be shifting to a model where recommended words will be precalculated rather than determined on a live basis for each user. Our changes will be integrated into the pre-existing recommendation engine, and we will ideally also be added to the aggregate user data in the form of a list of words that the user will learn.

Alternatives considered

We considered using advanced artificial intelligence or machine learning methods but decided to focus for now on a more feasible route. We also considered using live data but learned that this is the way the recommendation engine is currently done, and that Aleksey would prefer to move away from this.

Detailed description

In addition to the general changes discussed above, more specifically our changes will be affecting the Game Services API such that when developers query the API for games, our recommendation engine will be used in conjunction with user data to suggest a word. Furthermore, we will be updating the way in which words are suggested to users that are not signed in, in order to ensure that they are less random and follow a generic difficulty progression.

Backend

Our project will require storing new representations of words in the backend, as well as changing how the current recommendation works. The initial step in this will comprise of understanding how analytics are done in the current backend.

Security impact

We do not foresee security impacts beyond those created by the pre-existing recommendation engine, as we will largely be following the same architecture. The only potential security risk could be in storing pre-calculated words, but we would follow DTML policies for storing user data for this.

Deployment plan

Unclear so far, but we will aim to work primarily in Python, which should be supported with the current DTML stack. Any new additions will be communicated when known.

Cost estimate

So far, most costs will be managed by the team. For instance, we will be working on creating documentation for the pre-existing recommendation system to better understand it. In the future, however, we may need user statistics to determine how effective our changes are once they are implemented in production.