Towards a Word Sheriff 2.0: Lessons learnt and the road ahead

Galen Han[*]   **Mark Menezes**[*]   **Leah Halaseh**[*]
Jędrzej Stuczyński[*]   **Daniele Menara**[*]   **Sebastian Riedel**   **Pontus Stenetorp**
`{galen.han,mark.menezes,leah.al-halaseh}.14@ucl.ac.uk`
`{jedrzej.stuczynski,daniele.menara}.14@ucl.ac.uk`
`{s.riedel,p.stenetorp}@cs.ucl.ac.uk`
Department of Computer Science, University College London

Abstract

In this extended abstract we report on the ongoing work of improving a previously introduced Game With A Purpose. While the previous work argued that basing a Game With A Purpose on a well-established game would improve game design and long-term sustainability – we report several shortcomings with its implementation and how we strive to alleviate them based on current game trends and user feedback.

1   Introduction

Ever since the rise to dominance of statistical approaches to Natural Language Processing (NLP), the field has experienced an insatiable need for annotated data for both training and evaluating NLP systems. von Ahn and Dabbish (2004) introduced the idea of designing Games With A Purpose (GWAPs) in which players perform annotation tasks unwittingly as a part of gameplay while being motivated by their desire to be entertained. Since then, GWAPs have become a part of the NLP annotation creation toolbox and continue to have an impact on the field to this day (Otani et al., 2016).

Recently, Parasca et al. (2016) introduced the Word Sheriff, a language-based GWAP inspired by the popular television game show Pyramid, that proved largely successful in a limited release. They argued that the data collected by their GWAP would be useful for evaluating representations of linguistic knowledge in a communicative setting, as well as providing a resource to complement the shortcomings of representations trained using objectives inspired by the distributional hypothesis.

---

[*] Contributed equally to this work. Author order determined using a pseudorandom number generator.

In this extended abstract, we describe our efforts to produce a *Word Sheriff 2.0* that alleviates several shortcomings in their approach and implementation identified by prospective players.

2   Shortcomings with the 1.0 version

Through interviews with potential players we were able to highlight several key issues with the Word Sheriff version as proposed by Parasca et al. (2016). A common complaint was that the user interface lacked a social context, leading to other players appearing less human despite the intended social nature of the game. Several users attempted to access the game using mobile devices, which – while possible – proved to be less than ideal since the web application had not been designed with mobile devices in min. While the computer players – which substitutes for human players when not enough players are online – were functional, their behaviour was easily identified by human players as non-human and thus lead to frustration.

3   Ongoing and proposed directions

Currently, a team of five undergraduate students are using the observations in Section 2 to produce an improved version – of which a screenshot can be seen in Figure 1. Apart from visual improvements, we have also sought to make the game more responsive to reduce the amount of perceived delay a user experiences when waiting for input from other users.

While over the last decade a multitude of GWAPs have been proposed, only a few have stood the test of time. An early NLP GWAP would be von Ahn et al. (2006), which is no longer available online. To the best of our knowledge, the currently longest running successful NLP GWAP, which is still publicly accessible, is Phrase Detectives (Chamberlain et al., 2008). One key dif-
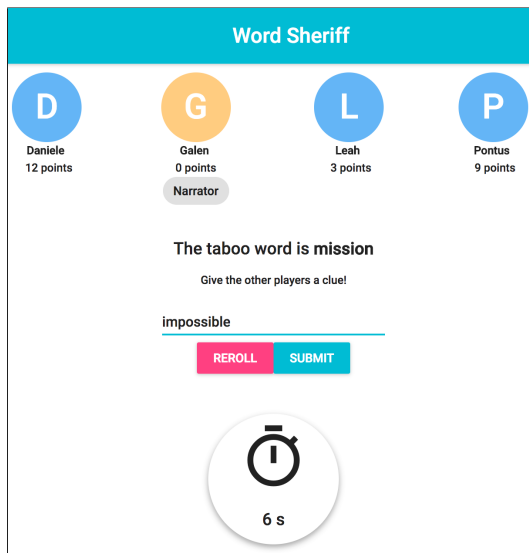
Figure 1: User Interface (UI) for the narrator view of our current working prototype. The new UI has a cleaner modern design to target an older demographic, as opposed to its predecessor that would most likely appeal to a significantly younger audience.

ference between the two is that the former lacks a social component, something which appears to be key not only for the longevity of GWAPs but also a majority of modern mobile games. Based on this observation and user feedback, our current prototype supports Facebook integration as well as anonymous logins. By integrating social media, interactions become more personal and we also plan to support a chat system with fixed responses and allowing players to invite friends for a game.

An ongoing trend is the gradual decline of PC and console games in favour of mobile games (De Prato et al., 2014). Unlike the previous version, our prototype is a responsive web application that allows us to target any device that supports a web browser. While not ideal for computationally demanding GWAPs, our GWAP only requires modest computational resources and can run on even modest mobile hardware.

Player-facing game bots – unlike hard AIs – must optimise for the enjoyability of a human player engaging with them. For example, pitting a master chess bot against a human beginner is unlikely to entice the human player to continue playing. Parasca et al. (2016) annotated 241 words as seeds for their bots, leading bots to having a distinct advantage since they had access to the target word list – making them appear super human. For example, if a clue word "fruit" was given, a human player could consider any fruit a possibility while the bots would have a choice between at most three fruits, making a wild guess for the bots more likely to succeed. We are resolving this issue by collecting additional seed words to diversify the vocabulary – at the time of writing we now cover a few thousand common English words.

In our ongoing efforts to improve the prototype we are planning a gathering of roughly sixty university students where in exchange for pizza for dinner they are asked to play the game for a limited period of time and fill out a short questionnaire. It is our belief that the first week of exposure when launching a GWAP is critical as disgruntled players may be unlikely to give a game a second chance, we thus want the user experience to be as good as possible before making our GWAP available to a wider audience and hope that this student feedback will prove helpful.

Acknowledgements

References

J. Chamberlain, M. Poesio, and U. Kruschwitz. 2008. Phrase detectives: A web-based collaborative annotation game. In *I-Semantics*.

G. De Prato, C. Feijóo, and J. P. Simon. 2014. Innovations in the video game industry: Changing global markets. *Digiworld Economic Journal*.

N. Otani, D. Kawahara, S. Kurohashi, N. Kaji, and M. Sassano. 2016. Large-scale acquisition of commonsense knowledge via a quiz game on a dialogue system. *OKBQA*.

I.-E. Parasca, A. L. Rauter, J. Roper, A. Rusinov, G. Bouchard, S. Riedel, and P. Stenetorp. 2016. Defining words with words: Beyond the distributional hypothesis. In *RepEval*.

L. von Ahn and L. Dabbish. 2004. Labeling images with a computer games. In *SIGCHI*.

L. von Ahn, M. Kedia, and M. Blum. 2006. Verbosity: a game for collecting common-sense facts. In *SIGCHI*.