NERDSearch : Using Familiar Gameplay Mechanics for Linguistic Annotation

Adwait Deshpande
Reverie Language Technologies
adwait.deshpande@reverieinc.com

Nischay Ghattamraju
Mendix
nischay.ghattamraju@mendix.com

Abstract

We present a linguistic annotation game with a purpose (GWAP) for a Named Entity Recognition task. The game employs the familiar "word search" puzzle mechanic for annotating low resource languages. This ensures that lesser time is spent on instructing players, allowing us to quickly dive into labeling new data. Another advantage is the ability to deploy the game in a commercial setting that provides a wider audience than most other serious games. We also use techniques such as inter-rater reliability of user input to ensure high quality and sanity of the data. This is a work in progress that, we hope, inspires other GWAP designers to use existing techniques to improve adaptability of linguistic annotation games.

1   Introduction

Humans can easily complete computationally challenging tasks using innate perceptual ability or commonsense reasoning. Games with a purpose (GWAPs) are designed with an intent to harness this human ability to assist computers with such tasks using games that entertain. Increasingly, GWAPs are being used for annotating large-scale linguistic data of varying complexity. This provides a fun alternative to a usually laborious process involving significant human computation hours. In this work, we are particularly inspired by Luis Von Ahn's work and his statement: "People play not because they are personally interested in solving an instance of a computational problem but because they wish to be entertained" [2]. We present Named Entity Recognition Data Search (*NERDSearch*), an adaptation of the highly popular "Word Search" game to annotate named entities in lower resource languages. The popularity of this puzzle game mechanic helps us do away with a long-drawn tutorial and explanation phase that accompanies most serious games. Another advantage is that we can build on an existing implementation, saving time on designing from scratch and focusing on data acquisition and analytics. The motivation to play *NERDSearch* is not primarily based on a desire to help the research community but from a desire to enjoy solving puzzles of increasing complexity.

2   Related Work

There have been several attempts to use popular word game mechanics and gamification techniques to annotate data. Wormingo [5] is one such game that was designed for the task of anaphoric annotation. By relying on a Motivation-Annotation paradigm, Wormingo tries to keep the user engaged by offering non-annotation related fun puzzles in between annotation levels. This novel paradigm and the usage of other gamification techniques in annotation GWAPs, try to tackle the common challenge of keeping a user engaged. However, our approach to tackle this challenge improves on previous attempts by choosing a game that is already popular. The game picked by us, "Word Search", is listed as one of the top games in the Word games category on Google Play. Our effort is similar to other

games like *Puzzle Racer* and *Ka-boom!* that use existing popular games like Mario Kart and Fruit Ninja respectively, but for image-sense association [4].

3 Game Design

3.1 Gameplay

NERDSearch uses the game design of word search with a few modifications. Our implementation is based on an open source Android implementation of the game[1]. For each new grid, the user is instructed to highlight words in the grid that belong to either of Name, Place or Organization categories. The user then hits finish (or "Zap", as we call it) on highlighting all the words that correspond to the category. The grid size starts from 5x5 and moves up to 12x12 as the player progress in the games. The game uses gold standard levels that contain expert annotated entries from the dataset. The user earns 50 points for each word entered in the gold standard annotated levels. Otherwise, the user earns 20 points plus a bonus card for the annotated words. A player earns points on the bonus card after a few days, once enough players annotate the same grid and a consensus is reached. These reward mechanics are inspired from "Wormingo" which uses eggs that hatch once the players response agrees with most other players [5].

3.2 Level Design

Each word grid of *NERDSearch* is constructed by scrambling a sentence from our dataset such that the characters of a word stay connected in the grid. Since the game is primarily aimed at annotating low resource Indian languages, the grid should also take care of Indic scripts. Most Indic scripts follow the "abugida" system, where a consonant and a vowel form one syllable unit [1]. While splitting words from abugida scripts, we ensure that the syllable units stay together for easy reading. The blank spaces in the grid are filled in with random characters or syllable units (see Figure 1). The initial few levels use gold standards from the dataset. These levels serve as tutorials and enable us to assess user reliability. As the game progresses further, the grid size increases, allowing players to annotate longer words and sentences. For every level we introduce a few words belonging to the current level category by randomly sampling from a large list of entities. The user would have to mark at least one of these words before proceeding further. In addition to this, we also inject a few gold standard levels (marked out as "Special Challenges") after every 10 games. This ensures user attention, keeping the data quality in check.

3.3 Data Collection

We present the users with an End-User License Agreement while registering to play on *NERDSearch* to obtain user consent for data collection. The collected data will be released as an open-source dataset for further research purposes.

4 Annotation Quality

We incorporate a reliability "meter" per user that aims to measure the weight that each user input carries towards the final annotations. These weights are calculated using triggers and metrics which are based on the user's success in the Gold Standard Levels and the position on the leaderboard. The objective is to boost the reliability meter of all players with high scores by assigning their data a higher weight when annotating a sentence. We consider a sentence to be successfully annotated once there are enough players in agreement. This can be achieved using "Krippendorff's Alpha" [3] which measures the degree of agreement between observers in an experiment. The advantage of using this measure is that it does not require same number of players rating each annotation.

5 Conclusion

Using an existing and familiar gameplay to design linguistic annotation games means greater portion of time is spent annotating unlabelled data. This also helps us hook in players and engage them in an

---

[1] https://github.com/abdularis/Word-Search-Game/

annotation task while masking the actual intent. By incorporating inter-rater and per user reliability we can build a robust crowdsourced dataset. In our future work we would like to explore additional gamification options that keep interested players engaged over a longer run. We believe that this approach towards NLP game design appeals to wider audiences thereby fast tracking the process of gathering data useful for NLP applications.

References

[1] Peter T Daniels and William Bright. *The world's writing systems*. Oxford University Press on Demand, 1996.

[2] Luis Von Ahn and Laura Dabbish. "Designing games with a purpose". In: *Communications of the ACM* 51.8 (2008), pp. 58–67.

[3] Klaus Krippendorff. "Computing Krippendorff's alpha-reliability". In: (2011).

[4] David Jurgens and Roberto Navigli. "It's all fun and games until someone annotates: Video games with a purpose for linguistic annotation". In: *Transactions of the Association for Computational Linguistics* 2 (2014), pp. 449–464.

[5] Doruk Kicikoglu et al. "Wormingo: a 'true gamification' approach to anaphoric annotation". In: *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 2019, pp. 1–7.
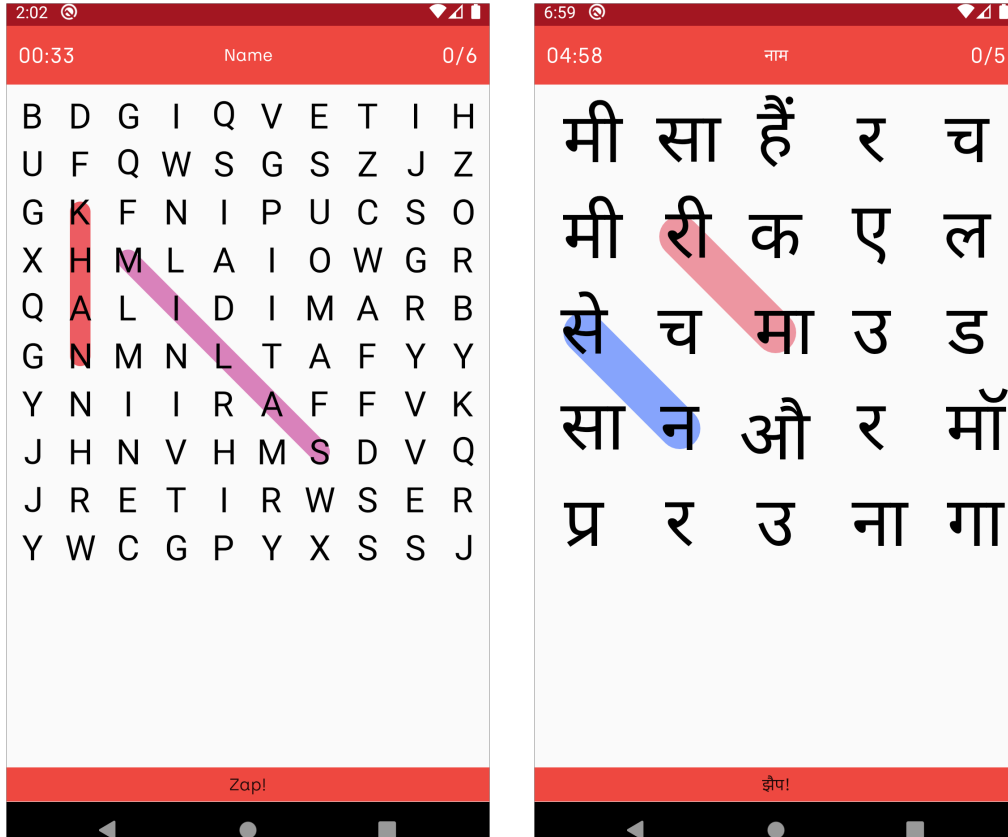
Appendix

Figure 1: Screenshots from the English (left) and the Hindi (right) versions of *NERDSearch*