GiveALink Tagging Game: An Incentive for Social Annotation

Li Weng*

Filippo Menczer

Center for Complex Networks and Systems Research
School of Informatics and Computing
Indiana University, Bloomington, IN, USA

ABSTRACT

Social tagging systems lead to inferred relationships among resources, tags and users from shared annotations in support of applications such as search, recommendation, and navigation. However, users share annotations largely for their own individual needs, such as bookmarking and involvement in online communities. Free tagging may also result in low-quality annotation data. In this paper, we introduce a tagging game designed as an incentive for users to share a large number of high-quality social annotation data while being entertained. Preliminary results suggest that playing the game leads users to classify resources into broad categories.

1. INTRODUCTION

In the GiveALink project (`cnets.indiana.edu/groups/nan/givealink`) we are working on several questions concerning the construction and applications of socially driven semantic annotation networks. One of them is the development of games as incentives for social participation in the process of tagging online resources. Here we discuss a tagging game that builds upon existing work in GiveALink. Previous research includes the design of effective similarity relationships in the semantic space of social tagging systems [4, 5], applications to page recommendation [6], exploratory navigation interfaces [1], bookmark management [9], and social spam detection [3]. We wish to explore the use of this prior work, especially similarity measures among Web pages and tags, and spam features, in our design of an effective game.

Social annotation is one of the ways in which Web users organize their bookmarks autonomously, by tagging them with keywords. By sharing their annotations online and viewing other people's annotations, users participate in a community. There are a number of successful Web applications of social tagging, such as `Delicious.com` and `Last.fm`. Most popular social annotation systems allow users to store, organize, and share resources.

There is little motivation for the majority of users to annotate many resources with sufficient numbers of accurate tags [2]. Motivations such the convenience of centralized bookmark management do not work for everyone, leading to low participation rate and short-time user involvement in tagging. The number of new pages that are posted per day to social annotation systems is small compared to the rate of growth of the Web [2]. The free tagging model also has some disadvantages: overly general tags; personal tags that are not meaningful to others; and tag abuse such as spam.

Our motivation for designing a game largely comes from the need to collect high-quality social annotation data, with the goal of supporting many applications (as discussed in § 2). The GiveALink tagging game proposed here aims to encourage users to tag Web pages, generating new triples by making the process of tagging enjoyable. The game algorithm and scoring mechanism are designed to favor specific and uncommon tags, and prevent cheating behaviors. The design objectives are explained in detail in § 3. We conclude the paper by describing our implementation of the game (§ 4) and outlining some preliminary analysis (§ 5).

*Corresponding author. Email: `weng@indiana.edu`

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
KDD-HCOMP '10, July 25, 2010, Washington, DC, USA
Copyright 2010 ACM 978-1-4503-0222-7 ...$10.00.

2. BACKGROUND

An *annotation* is defined as a *triple*, or tripartite relationship between a *user*, a *tag*, and a *resource*. In the remainder of this paper we will also refer to a *link* as a relationship between a resource and a tag, supported by the users annotating the resource with the tag.

The collection of social annotation data has been shown to be useful to improve social navigation [7], Web search [2], personalized recommendation services [6], and social links prediction [10]. Because of the important use of social annotation data to improve other Web services, we turn to games to help us generate more useful data.

The GiveALink tagging game leverages entertainment as an incentive for the contribution of useful annotations. The idea of *games with a purpose*, or crowdsourcing through play, was originally conceived by von Ahn [11]. The most famous instance is the ESP Game, asking players to label images. Games are based on the human desire to be entertained, and may help solve problems that are hard for computers [11]. *Serious games* are an older, similar idea in the context of training in military, education, and public health settings [12]. Games in `ClubBing.com` are designed to promote the Bing search engine. Players have to use Bing as a tool to collect hints for winning, thus trying functions that they might not have noticed before. Winners are awarded real products, which is a strong incentive for players.

Previous research in the GiveALink project lays the foundation of the tagging game. The Maximum Information Path (MIP) similarity measure has been shown to be efficient to compute and effective at capturing relationships among pages, tags, or users consistently with external benchmarks [4, 10]. We leverage MIP and social spam detection features [3] to set up the game, assist the user toward victory, and prevent cheating, as described below.

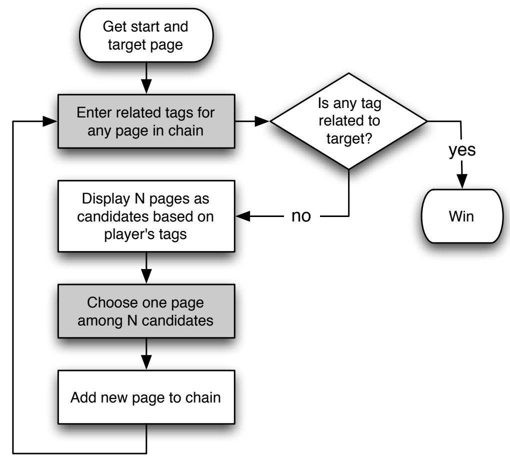

Figure 1: Flow chart of game logic. Gray boxes represent user actions, white boxes are system actions. Currently, $N = 3$.

3. GAME DESIGN

3.1 Basic Game Play

The basic idea of the game is to find a path (page chain) between two given pages by tagging. While players explore among online resources and try to build a path, new triples are generated.

The player is presented with a start page and a target page. The start page is the first node in the chain, and the player extends the chain by tagging any page inside the chain with one or more relevant tags. Each time the player enters tags for a page in the chain, the game displays a small set of pages based on those tags as candidates for the next node in the chain. The player is free to choose one of those candidates, and then select either this latest node or a previously tagged resource to be the next page to tag. The player wins the game when at least one of her latest tags are known to be relevant to the target. Therefore, once the player wins, a chain semantically connecting the start and target pages is successfully created. The logic is illustrated in Fig. 1.

3.2 Interface

The textfield for entering tags is placed between *target page* (top) and the *current page* (bottom), as shown in Fig. 2. The player needs to focus on tagging the current page and try to get it connected with the target. All pages in the chain are displayed in the cover flow style, so the player can easily switch between them. The url, title and thumbnail of each page are shown, so the player can try relevant tags based on that information or browsing the page by clicking on the thumbnail.

3.3 Mechanism Design

The data generated from the game is expected to be merged into the GiveALink triple store to decrease the sparseness of the semantics network of annotations and improve similarity calculations. Therefore, the game should be designed to strengthen relationships that appear weaker than they should be, or discover new relationships. At the same time, the game should be interesting enough to attract many players and should not be too hard to win.

The game mechanisms incorporate several design principles:

- Tags should be relevant to the current page in the chain; the more relevant they are, the more points the player can get.
- Links that exist in the GiveALink database are trusted, so they can be used as benchmarks.

Figure 2: Prototype game interface

- Links confirmed by many players can be trusted.
- New links are recorded until confirmed by other players.
- Once new links are confirmed by enough players, they can be added into GiveALink and treated as trustworthy later on.
- Candidate pages are selected based on the player's tags.
- Candidate pages should help the player approach the target, making the game easy to win.

The last principle requires an effective measure of similarity among pages. We use the Maximum Information Path measure from GiveALink [5]. MIP measures how similar a pair of tags or resources are. It has been implemented and integrated into the GiveALink system, and used as an important component in several applications, such as recommendation, bookmark management, and spam detection. The chain in the tagging game builds upon MIP similarities among pages, available via an API (`GiveALink.org/api_doc`); we select pages similar to the target to help the player.

Each tag for a given page (therefore each link) can be classified into one of three types:

1. **used**: tags that have been used *for this page* in GiveALink;
2. **suggested**: tags that have been suggested *for this page* by previous players, but not in GiveALink;
3. **new**: tags that have been used or suggested *for this page* by neither GiveALink users nor game players.

Because the **used** tags and the **suggested** tags are confirmed by either GiveALink or game players, they are more reliable. **New** tags can be relevant or irrelevant, so we need to wait for further evidence; they may turn into **suggested** tags when they are used for the same Web page more than once. **Suggested** tags may turn into **used** tags if they are confirmed by enough players. The players get more points with a larger proportion of **used** and **suggested** tags for a given page, as explained in § 3.4.

Once a player enters tags for the current page, a set of pages with any of these tags is selected based on their similarity to the target. The player should choose the next page that he believes is most similar to the target. The idea is that the similarity to the target increases as the chain is built, until the player can find a common tag between current and target pages, and win.

3.4 Scoring

Players are encouraged to provide more tags if possible, so more tags yield more points. The score of each tagging step is obtained by adding scores associated with each tag. Each tag is worth a number of points that depends on its specificity and the type of link it creates. Specific, trusted, and novel tags are most valued. The score for each tag t for the current page is given by two terms: $score(t) = \lambda(t) + \eta(t)$ where λ is a function of trustworthiness and novelty of t, and η is a measure of its *specificity:* $\eta(t) = \mu / \sum_{t' \neq t} \sigma(t', t)$ where σ is the MIP similarity between tags [5] and μ is a parameter. The denominator is a measure of *generality.* Tags that are similar to many other tags are general and therefore are worth fewer points.

According to the link classification, the trustworthiness and novelty functions are defined as follows:

- Discover trusted links: $\lambda(t_s) = \alpha \cdot f(t_s)$ where t_s is a **suggested** tag, f is the number of players who have suggested t_s for the current page, and α is a parameter. The player gets more points when more people have agreed with him in previous games.
- Find related tags: $\lambda(t_u) = \beta$ where t_u is a **used** tag for the current page and β is a parameter. We thus reward players who use relevant tags that have previously been used to annotate the same resource by GiveALink users.
- Add untrusted tags: $\lambda(t_n) = \gamma$ where t_n is a **new** tag for the current page and γ is a parameter.

Scores are added across all tags during each step of the game, and accumulated across the steps, yielding a cumulative score S. We set the parameters such that $\alpha > \beta > \gamma$ to prioritize based on the trustworthiness and novelty of links: among trusted links, novel ones (**suggested**) are more valued than known ones (**used**); untrusted (**new**) links are valued the least.

Additionally, winning the game yields a bonus that is larger if the player has won in fewer steps: $(S/\ell) + \delta$, where δ is a parameter and ℓ is the number of pages in the chain when the player wins.

3.5 Cheating

Cheating is an attempt to win the game without entering tags related to the current page. We define two types of cheating, each with its own solution.

Tag the Target. The first type of cheating is to tag the target page directly, without concern for the current one. This way the player can expect to win or at least get candidate pages similar to the target. To detect this behavior, the game determines whether the set of tags are relevant to the current page by computing the proportion of **suggested** or **used** tags. If most of the tags entered are **new**, the player is suspected of cheating. Specifically, a tagging step is deemed to be an instance of cheating if $\langle \lambda(t) \rangle_t - \gamma < \theta$, where $\langle \cdot \rangle_t$ denotes average over tags and θ is a threshold parameter.

Mixed Tags. The second type of cheating is to enter *some* tags that have been used for the target, but are irrelevant with respect to the current page. The player might use a mix of tags related to the current page and tags related to the target, such that the proportion of **used** or **suggested** tags is high enough to go undetected by the previous rule. We deal with mixed tags by noticing that tags aimed at different pages are likely unrelated to each other. We therefore use the $TagBlur$ measure, introduced for social spam detection [3], to catch this condition. For a set of tags T entered by the player, $TagBlur$ is computed from a measure of distance

Figure 3: Hints and score from the *Hintman* character

(dissimilarity) between tags, averaged across all the pairs of tags:

$$TagBlur(T) \propto \sum_{t_1 \neq t_2 \in T} \left(\frac{1}{\sigma(t_1, t_2) + \epsilon} - \frac{1}{1 + \epsilon} \right)$$

where ϵ is a small constant [3]. If the $TagBlur$ of the player's tags is lower than the $TagBlur$ of the current page in GiveALink, the player is suspected of cheating.

If no cheating is detected, the player can go forward toward the target (cf. § 3.3). If either cheating is detected, the player is not allowed to make progress; the candidate pages are selected among resources similar to the current page rather than the target.

3.6 Player-Friendly Design

A few game features are designed to make the game more fun. First, the target page and start page are not selected completely at random. To make a game easier to win, the player should be familiar with the two pages, and the two pages should be related. Players are registered users of GiveALink and may have already shared resources. If the player has annotated a sufficiently large number of pages (currently 50), we select one of these as the target, with some probability. Otherwise we select the page at random from the set of all GiveALink resources that have a minimum number of tags (5) and similar pages (5), and have been tagged by a minimum number of users not labeled as spammers (2). The start page is not chosen independently. We select it from the set of pages with small, nonzero similarity to the target. This way the start page is likely to be weakly related to the target — in fact they share at least one tag.

Spam can distract players and result in useless triples. Although the GiveALink API methods employ a spam filter, some spam can go undetected. The player can report a page as spam by simply clicking a button. Spam reports are stored, flagging suspicious resources and helping improve the accuracy of the classifier employed by the GiveALink spam detection system [3].

Finally, we set up a mechanism to display hints along with each player move, named *Hintman* (Fig. 3). Hintman helps the player review each move through different types of hints, telling him whether he is getting closer to or further away from the target, or whether his tags are related with the page in the previous step.

4. IMPLEMENTATION

When we combine the design principles, scoring rules, and anti-cheating mechanisms, we obtain the algorithm in Fig. 4. We have not yet implemented the specificity mechanism discussed in § 3.4.

The tagging game uses its own server and database for storing basic information on the players, triples generated during the games, scores, and other records.

The game implementation makes use of multiple GiveALink API methods. Initial pages are generated by the *url.generate* method. Then player u enters a set of tags T for page r, and the *url.getTags* method is used to judge whether tag $t \in T$ is **used** or not for r. If not, the triple (u, t, r) is added into the game database. A confirmed link (t, r) can be added into GiveALink via the *annotation.add* method. The similarity among pairs of tags in T is available through the *tag.similarity* method, necessary for the calcula-

```
Display start & target pages
UNTIL (win)
  Player picks start or next page as current page
  Player enters tags for current page
  FOREACH tag
    IF tag is suggested
      score += alpha * f(tag, current)
    ELSIF tag is used
      score += beta
    ELSE
      score += gamma
  IF [score >= (theta + gamma) * |tags|
      AND TagBlur(tags) < TagBlur(current)]
    IF win
      score += score/steps + delta
    ELSE
      Select pages similar to target as choices
  ELSE
    Select pages similar to current as choices
```

Figure 4: **Pseudocode of the game algorithm**

Table 1: **Basic game data statistics**

	Minimum	Average	Maximum
Score	10	194.4	775
Steps	1	2.7	9
Time	00:03	04:26	11:59

tion of $TagBlur$. For efficiency, $TagBlur$ is precomputed and retrievable via the API. With *tag.geturls* and *url.targetSimilarity* we can collect resources annotated with tags entered by the player, and rank them by their similarities to the target or current page.

The *cold-start* problem can negatively affect the performance of the game when there are only a few annotations available to build the chain. Therefore in our initial implementation we employ the Delicious API to obtain recommended tags for the target and accelerate the confirmation of new triples.

5. PRELIMINARY DATA ANALYSIS

The game prototype is at `GiveALink.org/giveagame`. Since we started testing within our research group, we have collected a small dataset consisting of 1,079 suggested and 265 confirmed triples. These annotations were generated by 19 players in the course of 97 games. Basic statistics are shown in Table 1.

In Table 2 we report, for the tags that appear in the most suggested triples, the *generality* (cf. § 3.4) and the percentage of GiveALink tags that are more specific. These preliminary results suggest that general tags, which can describe many resources, are easier to discover through the game; game players tend to classify pages in broad categories. This confirms our intuition that to build a chain one must move to a general category encompassing the start and target pages. At the same time we see the need to promote the use of specific tags as discussed in § 3.4.

Table 2: **Top 10 tags in suggested triples**

No.	Tag	Freq.	Generality	Percentage
1	online	24	35591.1	99.98%
2	italia	16	112.9	89.41%
3	web	13	15576.4	99.97%
4	blog	12	4903.5	99.88%
5	music	11	4516.3	99.87%
6	italy	10	478.1	96.87%
7	code	9	2406.4	99.60%
8	service	9	10880.4	99.96%
9	design	9	12613.6	99.97%
10	computer	8	3984.1	99.82%

6. CONCLUSION

Improving participation in social tagging systems is an important goal, since user-generated annotations are valuable. We have described the design of a game as an incentive for quality tagging, to relieve the problem of sparse annotation networks and support other Web applications.

Our design principles lead to basic rules to make the game enjoyable. Novel tags are rewarded in the game, but they need to be confirmed by multiple players. When a user is playing, she needs to use fresh tags and guess what might be used by other players. Therefore, the data originated by the game provides new triples with confirmed relatedness between tags and resources in each link.

The game design and implementation are under continuous development and testing. We plan to incorporate specificity into the scoring mechanism to reward specific or fresh tags. During the current internal test, both the mechanisms and interface are being refined. We are also tuning the values of the various scoring parameters. Once the game passes this internal test, it will be advertised to the public to generate more annotation data.

Analysis of the annotation data from the game and its comparison with existing social tagging systems will allow us to explore some interesting questions. First, we will evaluate the quality of game data: whether tags are more or less specific, and more or less relevant, compared to those from bookmarks. Second, we will track changes in the structure of the folksonomy built from the relationships between tags, users, and resources when adding the game data. For example, we want to see if the game makes the similarity networks among tags and resources more "metric" [8]. Finally, we plan to compare the speed with which triples are generated via the game versus traditional bookmarking.

Acknowledgements. We thank C. Cattuto, R. Schifanella, and NaN members (`cnets.indiana.edu/groups/nan`) at the IU School of Informatics and Computing for helpful discussions and participation in testing. This work is funded by NSF award IIS-0811994.

7. REFERENCES

[1] J. Donaldson, M. Conover, B. Markines, H. Roinestad, and F. Menczer. Visualizing social links in exploratory search. In *Proc. 19th Conf. on Hypertext and Hypermedia (HT)*, 2008.

[2] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? In *Proc. 1st WSDM Conf.*, 2008.

[3] B. Markines, C. Cattuto, and F. Menczer. Social spam detection. In *Proc. 5th AIRWeb Workshop*, 2009.

[4] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. Evaluating similarity measures for emergent semantics of social tagging. In *Proc. 18th Intl. WWW Conf.*, 2009.

[5] B. Markines and F. Menczer. A scalable, collaborative similarity measure for social annotation systems. In *Proc. 20th Conference on Hypertext and Hypermedia (HT)*, 2009.

[6] B. Markines, L. Stoilova, and F. Menczer. Social bookmarks for collaborative search and recommendation. In *Proc. 21st Natl. Conf. on Artificial Intelligence (AAAI)*, pages 1375–1380, 2006.

[7] D. Millen and J. Feinberg. Using social tagging to improve social navigation. In *Adaptive Hypertext Workshop on Social navigation and community-based adaptation*, 2006.

[8] L. M. Rocha. Semi-metric behavior in document networks and its application to recommendation systems. In *Soft Computing Agents: A New Perspective for Dynamic Information Systems*. 2002.

[9] H. Roinestad, J. Burgoon, B. Markines, and F. Menczer. Incentives for social annotation. In *Proc. 32nd SIGIR Conf.*, page 838, 2009.

[10] R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer. Folks in folksonomies: Social link prediction from shared metadata. In *Proc. 3rd WSDM Conf.*, 2010.

[11] L. von Ahn. Games with a purpose. *IEEE Comp.*, 39(6):92–94, 2006.

[12] M. Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, 2005.