BetterRelations: Using a Game to Rate Linked Data Triples

Jörn Hees[1,2], Thomas Roth-Berghofer[2,3], Ralf Biedert[2],
Benjamin Adrian[2], and Andreas Dengel[1,2]

[1] Computer Science Department, University of Kaiserslautern, Germany
[2] Knowledge Management Department, DFKI GmbH, Kaiserslautern, Germany
[3] Institute of Computer Science, University of Hildesheim, Germany
{firstname.lastname}@dfki.de

Abstract. While associations between concepts in our memory have different strengths, explicit strengths of links (edge weights) are missing in Linked Data. In order to build a collection of such edge weights, we created a web-game prototype that ranks triples by importance. In this paper we briefly describe the game, Linked Data preprocessing aspects, and the promising results of an evaluation of the game.

1 Introduction

Since its introduction in 2001 the Semantic Web [1] has gained much attention. In recent years, especially the Linking Open Data (LOD) project contributed many large, interlinked and publicly accessible RDF datasets, generating one of the world's largest, decentralized knowledge bases. The accumulated amount of Linked Data has many applications and can already be used to answer structured questions (e.g., the DBpedia [2] dataset can easily be used to compile a list of musicians who were born in Berlin).

Currently it is impossible to rank result sets—not even those of simplistic (descriptive) queries—by importance as considered by an average human. For example, asked to describe ("What/Who is ...?") *Facebook*, nearly all humans will explain that it *is an online social network*, but only few will tell us that *Chris Hughes is one of its co-founders*.[4] In the remainder of this paper, we will hence call the fact "Facebook has subject online social networking" to be more *important* than "Facebook has key person Chris Hughes".

In order to overcome the knowledge acquisition bottleneck, which involves the manual generation of a dataset of such explicit importance ratings for many facts, we sketched the idea for a web-game in [3]. In this paper we present our experiences gathered from a web-game prototype called *BetterRelations* following the "Games With A Purpose" approach by von Ahn and Dabbish [4].

[4] In this paper we focus on an "average human's" view, leaving the application of user and context models to future work.

2 Related Work

In terms of game design, BetterRelations is related to *Matchin* [5], which confronts both players with two pictures (taken from the WWW), asking them which one they prefer. In contrast, BetterRelations presents two textual facts about one topic to its players. Whereas Matchin returns a globally ranked list of images, BetterRelations creates a ranking for each topic and its related facts. In order to avoid forced decisions in cases of unknown or noisy facts, the GUI had to be extended, also causing the need to modify Matchin's reward function in order to counter obvious cheating strategies.

OntoGame [6] was the first and most prominent game with a purpose focusing on Linked Data. Nevertheless, it collects another type of information than BetterRelations: Players are asked to decide if a Wikipedia topic is a class or an instance, aiming at creating a taxonomy of Wikipedia.

WhoKnows? [7], a *single player* game, judges whether an existing Linked Data triple is known by testing players with (amongst others) a multiple choice test or a hangman game. In contrast to our approach, WhoKnows only uses a limited fraction of the DBpedia dataset and excludes triples not matched by a predefined domain ontology in a preprocessing step. This greatly reduces noise issues, but eliminates the possibility to collect user feedback about triple qualities and problems in the extraction process. Also, WhoKnows intends to rank triples by degree of familiarity. However, the used measurement only relies on the ratio of correctly recognized facts divided by number of times a fact was tested. The quality of this ratio is doubtful as it does not distinguish whether a fact has been tested few or many times.

3 The Game

A straightforward approach to collect association strengths for Linked Data triples is this: First, we select a Linked Data resource of interest (e.g., `dbpedia:Facebook` or `dbpedia:Wiki`). We call this a *topic of interest* or simply *topic*. We then show randomly shuffled lists of all related triples to test persons and ask them to order the triples by decreasing importance. In the context of this work, given a topic, we define *related triples* to be the collection of (subject, predicate, object)-triples where the topic is the subject.[5]

The aforementioned approach suffers from the problem that the outcome of each of these experiments, which is a user centric ranking, is not only highly subjective, but sometimes even unstable for one person over time. In order to overcome difficulties for humans when sorting lengthy lists, we could ask for the atomic relative comparisons of two facts about one topic and then use an objective rating algorithm to generate an absolute ranking of the topic's related facts. This leads us to the idea behind BetterRelations.

[5] Extending the list by triples where the topic is the object (incoming links) typically imports a large number of unimportant facts for the topic (e.g., in Wikipedia and thus in DBpedia one would expect to learn about Facebook by visiting the page about it, not by reading through all the pages linking to its page).

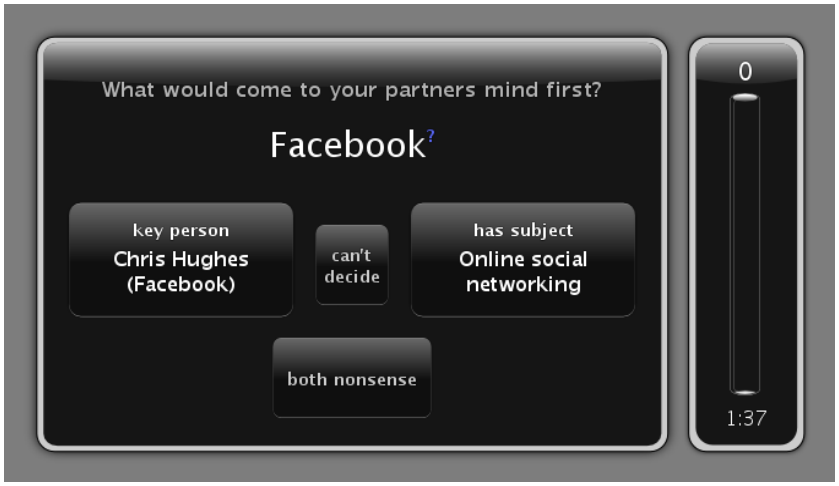

Fig. 1. In a game round, choosing phase.

3.1 BetterRelations

BetterRelations[6] is a symmetric two player output (decision) agreement game in terms of von Ahn and Dabbish's design principles for Games With A Purpose [4]:

A player starting to play the game is randomly matched with some other player for a predefined timespan (e.g., 2 minutes). In every round (see Figure 1) both players are presented with a *topic*, which actually is a Linked Data resource's symbol (e.g., *Facebook*, the symbol for `dbpedia:Facebook`), and two *items*, which are symbolic forms of facts about the topic (e.g., *key person Chris Hughes (Facebook)* and *has subject Online social networking*).

Both players are asked to select the fact that their partner will have thought of first. In case a player does not know the topic, a quick info can be requested by clicking on the question mark appended to the topic. Doing so will internally mark the player's decision as influenced and the partner's as unvalidated. To decide, each player can either click on the more important fact's button or on two additional buttons in case the player can't decide between the alternatives or thinks that both alternatives are nonsense / noise.

On the server side the game records a large amount of relative decisions between pairs of items, filtered by a partner and uses them to upgrade ratings in case of agreements. Internally, BetterRelations uses a TrueSkill [9] based algorithm to update fact ratings after each agreement, selects next fact pairs for a topic in a way to minimize the overall needed amount of decisions and stops sorting lists with n facts after $n \cdot \log_2(n)$ updates, determined to be a good threshold by simulations.

After rewarding the players with points, the next round starts until the game runs out of time. The next topic is chosen by selecting the topic least often

[6] BetterRelations can be played online: `http://lodgames.kl.dfki.de`

played by both players from a list of topics currently opened for playing, which is based on the topmost accessed Wikipedia articles. In the end, both players see a summary of their performance showing the amount of points gained in this game, the longest streak and their total game score in BetterRelations.

In case no partner can be found or the partner leaves the Game, BetterRelations also provides a single player mode.

3.2 Game Data Acquisition and Preprocessing

In order to provide players with popular topics, BetterRelations selects topics (URI references, e.g., `http://dbpedia.org/resource/Facebook`) corresponding to the most often accessed Wikipedia pages[7].

Each time the game needs a new game topic and its related triples (e.g., because an existing topic's facts were sorted), it loads the corresponding triples for the next topmost Wikipedia topic from a local DBpedia mirror, which also was pre-loaded with standard vocabularies such as `rdf`, `rdfs`, `foaf`.

As showing URIs to the end-users is of limited use, the users will always see `rdfs:labels` of such references. Triples having the same labels are merged from a game's point of view and such with missing labels for predicate or object excluded from the game.

Finally labels and corresponding triples are excluded, which (due to long string length) don't fit into the game's window, end with suspicious file endings (e.g., .jpeg) or which have an object label equal to the topic's label ("Facebook label Facebook").

4 Evaluation

BetterRelations was tested in an 18 day period in January 2011. In this time 1041 games were played by 359 users, resulting in over 4700 matches within an overall playtime of 42 human hours. From this we can estimate an *average lifetime play* of 7 minutes per player, a *throughput* of 112 matches per human hour of gaming, and an *expected contribution* of 13 matches per player.[8] Furthermore, with our current approach, we can estimate, that in order to sort the facts known about the top 1000 Wikipedia topics we would need about $313K$ matches or $23.9K$ players, so 24 players per Wikipedia topic.

We also compared the resulting ordering of facts with a manually created gold standard and found out that the rankings generated by BetterRelations can compete with those generated by human beings: In half of the cases (6/12) our approach won against the average single human's error. In three more it was approximately equal.

[7] Stats aggregated from raw access logs, available at `http://dom.as/wikistats/`

[8] Throughput, average lifetime play and expected contribution as in [4].

5 Conclusion

In this paper we presented results from implementing and testing BetterRelations, a game with a purpose which rates Linked Data triples by importance.

Our evaluation shows very promising results in terms of the desired and achieved high quality of the generated collection of importance ratings. However, the low average lifetime play indicates a problem with the game's fun factor. Based on a questionnaire we identified the high amount of noise in the underlying Linked Data triples to be the main problem (i.e., nonsense, unknown, and irrelevant facts).

As even slight improvements of the low average lifetime play could already drastically reduce the number of players needed to sort the facts known about a popular Wikipedia topic, our future work will focus on ways to reduce the amount of noise included in BetterRelations and other ways to increase the player's fun, such as including user accounts and high scores. We also plan to provide the game's output (ranked lists with rating scores) as Linked Data, allowing others to rank result sets of queries by importance for humans.

This work was financed in part by the University of Kaiserslautern PhD scholarship program and the BMBF project Perspecting (Grant 01IW08002).

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American **284**(5) (May 2001) 34–43
2. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - A crystallization point for the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web **7**(3) (September 2009) 154–165
3. Hees, J., Roth-Berghofer, T., Dengel, A.: Linked Data Games: Simulating Human Association with Linked Data. In Atzmüller, M., Benz, D., Hotho, A., Stumme, G., eds.: LWA 2010, Kassel, Germany (2010)
4. von Ahn, L., Dabbish, L.: Designing games with a purpose. Communications of the ACM **51**(8) (August 2008) 58–67
5. Hacker, S., von Ahn, L.: Matchin: Eliciting User Preferences with an Online Game. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, ACM (2009) 1207–1216
6. Siorpaes, K., Hepp, M.: OntoGame: Towards Overcoming the Incentive Bottleneck in Ontology Building. In: Proc. of the 3rd International IFIP Workshop On Semantic Web & Web Semantics (SWWS), Vilamoura, Portugal, Springer, LNCS (2007)
7. Kny, E., Kölle, S., Töpper, G., Wittmers, E.: WhoKnows? (October 2010)
8. Singh, P.: The Open Mind Common Sense Project (2002)
9. Herbrich, R., Minka, T., Graepel, T.: TrueSkill(TM): A Bayesian Skill Rating System. In Schölkopf, B., Platt, J.Hoffmann, T., eds.: Advances in Neural Information Processing Systems. Volume 19., Cambridge, MA, USA, MIT Press (2007) 569–576