Evaluating Entity Summarization Using a Game-Based Ground Truth

Andreas Thalhammer,[1] Magnus Knuth,[2] and Harald Sack[2]

[1] University of Innsbruck, Technikerstr. 21a, A-6020 Innsbruck
andreas.thalhammer@sti2.at
[2] Hasso Plattner Institute Potsdam, Prof.-Dr.-Helmert-Str. 2–3, D-14482 Potsdam
{magnus.knuth, harald.sack}@hpi.uni-potsdam.de

Abstract. In recent years, strategies for Linked Data consumption have caught attention in Semantic Web research. For direct consumption by users, Linked Data mashups, interfaces, and visualizations have become a popular research area. Many approaches in this field aim to make Linked Data interaction more user friendly to improve its accessibility for non-technical users. A subtask for Linked Data interfaces is to present entities and their properties in a concise form. In general, these summaries take individual attributes and sometimes user contexts and preferences into account. But the objective evaluation of the quality of such summaries is an expensive task. In this paper we introduce a game-based approach aiming to establish a ground truth for the evaluation of entity summarization. We exemplify the applicability of the approach by evaluating two recent summarization approaches.

Keywords: entity summarization, property ranking, evaluation, linked data, games with a purpose

1 Introduction

The main idea of the Semantic Web is to make implicit knowledge explicit and machine processable. However, machines that process knowledge are not a dead end. In fact, after processing the returned results are either consumed by another machine or by human users. In this paper, we focus on the latter: the consumption of machine processed data by human users. A lot of efforts in the Semantic Web currently focus on Linked Data interfaces and Linked Data visualization. As for the former, most interfaces have been developed by the Linked Data community and usually show all information (usually as property-value pairs) that is available for an entity (e. g. Pubby[3], Ontowiki[4], etc.) and leave it to the user to decide which of the information is important or of interest. In May 2012, Google[5] introduced its "Knowledge Graph" (GKG), which produces summaries for Linked Data entities. While it is not the first approach to rank properties or

[3] Pubby – http://www4.wiwiss.fu-berlin.de/pubby/
[4] Ontowiki – http://ontowiki.net/
[5] Google – http://google.com/

features of Linked Open Data according to their relevance [9,11,3] the uptake by industry certainly gives incentives for further investigation in this subject. This has to be considered in line with the fact that Google processed 87.8 billion queries in December 2009 [4] which makes roughly 2.8 billion queries per day. Keeping the huge number of daily searches in mind, it was an interesting move by Google to devote a big part of its result pages to the GKG summaries. Having an average of 192 facts attached to an entity [3], producing a concise summary that is shaped to an entity's individual characteristics states an interesting research problem.

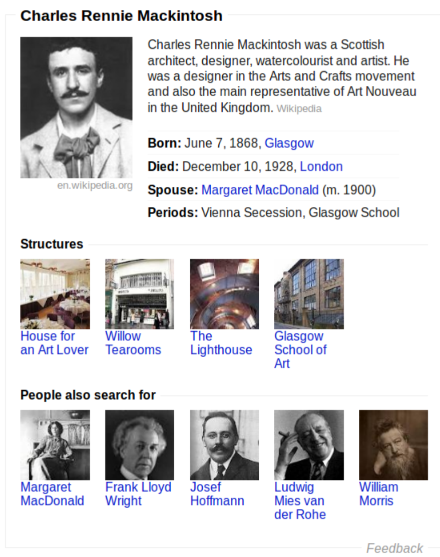
In this paper we will discuss current developments in Linked Data entity summarization and fact ranking as well as the need for a gold standard in form of a reference dataset which makes evaluation results comparable. We introduce a novel application of games with a purpose (GWAPs) that enables us to produce a gold standard for the evaluation of entity summarization. We demonstrate the applicability of the derived data by evaluating two different systems that utilize user data for producing summaries (one of which is GKG). In the course of our explanations we will emphasize on the complete and correct description of our test settings and stress that all data (that does not violate the privacy of our users) is made publicly available.

The remainder of this paper is structured as follows: Section 2 gives a description of the state-of-the-art in Linked Data entity summarization including the Google Knowledge Graph. In Section 3 the processed data sets, the quiz game and the evaluated systems are explained in detail, while Section 4 reports the achieved results. Section 5 concludes the paper with a brief summary and an outlook on future work.

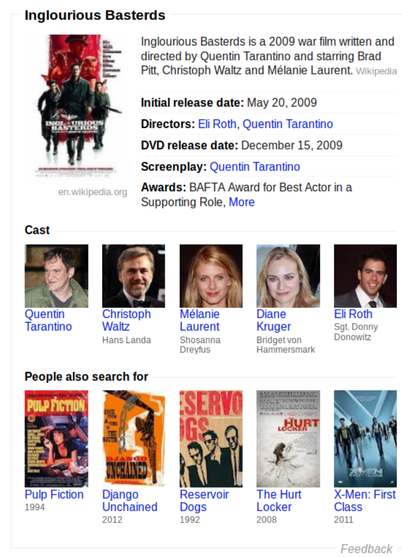
2 Background

In recent years, four approaches to Linked Data entity summarization have emerged including the one adopted by GKG. In the following, we will discuss all of those approaches and - in addition - present methods used for evaluating text summarization.

Google has introduced the "Knowledge Graph" in May 2012 [8]. The main idea is to enrich search results with information about named entities. In case of ambiguous queries, such as "lion king" (currently a musical and a film are returned), Google lists also different possibilities. Two examples for GKG summaries are shown in Fig. 1. Google's summaries are usually structured as follows: After presenting the name of the entity and an attached plot (usually taken from Wikipedia) next to a picture, up to five "main facts" are listed. These facts differ heavily between entities of different RDF types but also – to a certain extent – between entities of the same RDF type. After that, for certain RDF types like architects or movies, domain-specific attributes such as *'Structures'* (architects) or *'Cast'* (movies) are presented. For those, Google also defines a ranking e. g. from left to right for the *'Cast'* lists. In addition, a range of related entities is dis-



(a) GKG: architect and designer Charles Rennie Mackintosh.



(b) GKG: movie titled "Inglourious Basterds".

Fig. 1: Examples for GKG summaries (Source: `http://google.com/`).

played (Google introduces this list with *'People also search for'*). In their blog, Google developers describe summaries as one of "three main ways" to enhance search results with GKG information [8]. To automatically generate summaries, Google utilizes the data of their users, i. e. the queries, "[...] and study in aggregate what they've been asking Google about each item" [8]. We assume that these queries are in most cases "subject+predicate" queries, such as "lake garda depth", or "subject+object" queries such as "the shining stanley kubrick". In some cases also "subject+predicate+object" queries might make sense such as "jk rowling write harry potter"[6]. It is worth mentioning that using queries for determining the users' average interest in facts also has some pitfalls. For example, the query "inglourious basterds quentin tarantino" (querying for a movie and one of its directors) not only boosts the 'directed by' property but also the 'starring' property for the movie's relation to the person Quentin Tarantino. Unfortunately, this leads to the situation that the main actor (namely Brad Pitt) is not mentioned in the cast list while the director – who is known for taking minor roles in his movies and is doing so in this particular one – takes his position (see Fig. 1b).

Thalhammer et al. [9] explain how entity neighborhoods, derived by mining usage data, may help to discover relevant features of movie entities. The authors outline their idea that implicit or explicit feedback by users, provided by

[6] In fact, this query was suggested by Google Instant (`http://www.google.com/insidesearch/features/instant/about.html`)

consuming or rating entities, may help to discover important semantic relationships between entities. Having established the neighborhood of an entity with methods adopted from item-based collaborative filtering [7], the frequency of a feature that is shared with its neighbors is likely to give an indication about the feature's importance for the entity. A TF-IDF-related weighting scheme is also adopted as some features are generally very common (e. g., provenance statements). Unfortunately, the authors do not provide an evaluation of their system and only provide some preliminary results. In the later sections, we will refer to this approach as UBES (usage-based entity summarization).

The term of "entity summarization" was initially introduced by [3]. According to the authors, entity summarization is the task of identifying features that "not just represent the main themes of the original data, but rather, can best identify the underlying entity" [3]. We do not fully agree with this definition. Rather than selecting features that unambiguously identify an entity, we suggest to select features that are most interesting to present to a user. Of course, for many entities there is a significant overlap between the features that best identify an entity and features that are most interesting for the users. As a further contribution, the authors introduce the term "feature" as a property-value pair. The approach presented in [3] applies a "goal directed surfer" which is an adapted version of the random surfer model that is also used in the PageRank algorithm. The main idea is to combine informativeness and relatedness for the ranking of features. In the conclusion of [3], the authors state that "user-specific notion of informativeness [...] could be implemented by leveraging user profiles or feedback" in order to mitigate the problem of presenting summaries that help domain experts but are not as useful for average users. The presented approach does not utilize user or usage data in order to provide summaries. However, this information could be given implicitly by the frequency of in and out links.

Waitelonis and Sack explain how exploratory search can be realized by applying heuristics that suggest related entities [11]. Assume that a user is currently browsing the current US president's Linked Open Data description. Attached to the president's URI are properties such as `dbpedia-owl:residence`, `dbpprop:predecessor`, or `dbpedia-owl:party`. Obviously, these links are useful to show in the context of exploratory search. However, as there are more than 200 facts attached to the entity, the authors propose to filter out less important associations (i. e., provide summaries). To achieve this, they propose and evaluate eleven different heuristics and various selected combinations for ranking properties. These heuristics rely on patterns that are inherent to the graph, i. e. they do not consider usage or user data. The authors conduct a quantitative evaluation in order to find out which heuristic or combination performs best. The results show that some heuristics, such as the Wikilink and Backlink-based ones, provide high recall while Frequency and Same-RDF-type-based heuristics enable high precision. Trials with blending also showed that either precision or recall can be kept at a significant high level, but not both at the same time. Like in the approach of GKG, the predicate and the object are decoupled. While the introduced heuristics address the predicates, the data gathering for the evaluation focuses on the objects. As exemplified above, this leaves space for ambiguity. In the discussion, the authors argue that summaries should be considered in a specific context (i. e., "what is the search task?") and therefore quantitative measures might not provide the right means to evaluate property rankings.

[3] and [11] provide evaluations of their approaches. Both provide a quantitative as well as a qualitative evaluation. In the quantitative evaluation, both approaches base their evaluation on DBpedia[7] excerpts comprised of 115 [11] and 149 [3] entities. These entities were given to a sufficient amount of users in order to establish a ground truth with human created summaries. To the best of our knowledge, the results of these efforts are not publicly available.

In the field of automatic text summarization, [1] discusses two possible ways for evaluating summaries: *human assessments* and *proximity to a gold standard*. Thus, in this area, not only a gold standard had to be created but also a way to measure closeness to such a reference. As entity summarization deals with structured data only, such proximity measures are not needed: to measure the similarity between a summary and a ground truth, we can make use of classic information retrieval methods such as precision/recall, Kendall's τ and Spearman's rank correlation coefficient.

3 Evaluating Entity Summarization

We attempt to create a ground truth for the task of entity summarization by utilizing data gained from a game with a purpose. We exemplify our approach in the domain of movies. Thus, our research hypothesis is as follows:

> *A game-based ground truth is suitable for evaluating the performance of summarization approaches in the movie domain.*

Our assumption is that implemented approaches that provide summaries should perform significantly better than randomly generated summaries when measuring the correlation to the established ground truth. It is important to note that the relevance of facts for the task of summarization will be evaluated on the entity level. This means that the same properties, objects, or even property-value pairs are of different importance for different subjects. As a matter of fact, the importance of facts for an entity might vary given different contexts and summarization purposes. However, summarization also involves a certain level of pragmatics, i. e. trying to capture the common sense to address as many users as possible.

In the following we detail the restraints for the chosen domain, the design of the quiz game, the interpretation of the gained data, and the experimental setup for the evaluated systems.

[7] DBpedia - `http://dbpedia.org/`

Listing 1: Property chain for defining a "hasActor" property.

```
<http://some-name.space/hasActor>
<http://www.w3.org/2002/07/owl#propertyChainAxiom> (
        <http://rdf.freebase.com/ns/film.film.starring>
        <http://rdf.freebase.com/ns/film.performance.actor> ).
```

3.1 Employed Dataset

In our evaluation, we focus on movie entities taken from Freebase[8]. This dataset contains a large amount of openly available data and – in contrast to DBpedia and the Linked Movie Database (LinkedMDB)[9] – very detailed and well curated information. Large parts of this dataset are also used by Google for its summaries [8]. For the evaluation, we have randomly selected 60 movies of the IMDb Top 250 movies[10] and derived the Freebase identifiers by querying Freebase for the property `imdb_id`. With facts about 250 movies, it is difficult to achieve the mandatory number of game participants for sufficient coverage. Therefore, we have restricted the number of movies to 60. We have downloaded RDF descriptions of the movies and stored them in an OWLIM[11] triple store with OWL2 RL[12] reasoning enabled. This enables us to connect properties (such as actors) that are linked via reification (such as the 'film-actor-role' relationship) directly with property chain reasoning. An example for creating such an axiom is provided in Listing 1. We have created such direct links for actors, role names, achieved awards, budgets, and running times. As a matter of fact, not all properties are useful to be questioned in a game. Therefore, we make use of a white list. The list of selected movies, the used property chain rules as well as the property white list are available online (cf. Sec. 4.3).

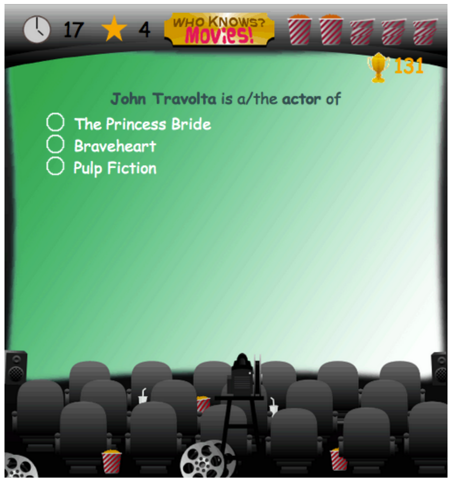
3.2 *WhoKnows?Movies!* – Concept and Realization

We developed *WhoKnows?Movies!* [10], an online quiz game in the style of *'Who Wants to Be a Millionaire?'*, to obtain a ground truth for the relevance of facts. The principle of the game is to present multiple choice questions to the player that have been generated out of the respective facts about a number of entities. In this case we limited the dataset as described in Sec. 3.1. The players can score points by answering the question correctly within a limited period of time and lose points and lives when giving no or wrong answers.

As an example, Fig. 2 shows the question *'John Travolta is the actor of ...?'* with the expected answer *'Pulp Fiction'*, which originates from the triple

```
fb:en.pulp_fiction test:hasActor fb:en.john_travolta .
```

[8] Freebase – `http://www.freebase.com/`
[9] LinkedMDB – `http://www.linkedmdb.org/`
[10] IMDB Top 250 – `http://www.imdb.com/chart/top`
[11] OWLIM – `http://www.ontotext.com/owlim`
[12] OWL2 RL – `http://www.w3.org/TR/owl2-profiles/#OWL_2_RL`



Subject	Property	Object
Pulp Fiction	**actor**	**John Travolta**
	actor	Uma Thurman
	actor	...
Braveheart	actor	Mel Gibson
	actor	Sophie Marceau
	actor	...
The Princess Bride	actor	Robin Wright
	actor	Annie Dyson
	actor	...

Fig. 2: Screenshot and triples used to generate a One-To-One question.

and is composed by turning the triple's order upside down: *'Object is the property of: subject1, subject2, subject3...'*. The remaining options are selected from entities that apply the same property at least once, but are not linked to the object of the question. In this way we assure that only wrong answers are presented as alternative choices. There are two variants of questions: *One-To-One* where exactly one answer is correct and *One-To-N* where one or more answers are correct.

When the player answers a question correctly he scores points and steps one level up, while incorrect answer will be penalized by loosing points and one live. The earned score depends on the correctness of the answer and the time needed for giving the answer. With growing level the number of options raises, so correct answers are getting harder to guess. It has to be noted that the probability for a fact to appear in a question with many or few choices is equal for all facts. This ensures that the result is not skewed, for example by putting some facts in questions with two choices only. When submitting an answer, the user receives immediate feedback about the correctness of his answer in the result panel, where all choices are shown once again and the expected answer is highlighted. Given answers will be logged for later traceability and the triple's statistics are updated accordingly. The game finishes when the player lost all of his five lives.

Applying the white list described in Sec. 3.1, 2,829 distinct triples were produced in total. For each triple a set of false answers is preprocessed and stored to a database. When generating a question for a specific triple, a number of false subjects is randomly selected from this set.

3.3 What are *interesting facts*?

The answer patterns of quiz games can tell a lot about what is generally interesting about an entity and what is not. One of the questions in the quiz game of Sec. 3.2 is *'What is the prequel of Star Wars Episode VI?'* with one of the answer options being *'Star Wars Episode V'*. Of course, most of the players were right on this question. On the other hand fewer players were right on the question whether *'Hannibal rising'* is a prequel of *'The silence of the lambs'*. The idea of a good general[13] summary is to show facts that are common sense but not too common. This is related to Luhn's ideas about "significance" of words and sentences for the task of automatically creating literature abstracts [6]. Transferring the idea about "resolving power of words" to the answer patterns of the quiz game, we can state that neither the most known nor the most unknown facts are relevant for a good summary, it is the part between those two. Unfortunately, we have not been able to accumulate enough data to provide a good estimation for fine grained upper and lower cut-off levels. Therefore, in Sec. 4 we measure the relevance correlation with a pure top-down ranking.

In addition, there might be questions, where not knowing the right answer for a given fact does not necessarily mean that this fact does not have any importance. For our movie quiz game, participants are also asked for actors of a given movie. First of all, Freebase data does not distinguish between main actors and supporting actors. Thus, the property actor might not be in general considered as an important property, because most people do not know many of the supporting actors. Furthermore, an actor might play a very important role in a movie, but the game players do not know his name, because they only remember the face of the actor from the movie. The same holds for music played in the movie, where the participants might not know the title but are familiar with the tune. Thus, for future use, also the use of multimedia data should be considered to support the text-based questions of the quiz game.

3.4 Evaluated Systems

We exemplify the introduced evaluation approach to the summaries produced by GKG [8] and UBES [9]. For both approaches the additional background data stems from user behavior or actions. In addition, the rationale of both systems is to present useful information to the end users in a concise way. These similarities guarantee a comparison on a fairly equal level. In this section, we will detail the experimental setup and the data acquisition[14].

Usage-based Entity Summarization (UBES)
In addition to Freebase, the UBES system utilizes the usage data of the HetRec2011 MovieLens2k dataset [2]. With a simple heuristic based on IMDb identifiers, more than 10,000 out of 10,197 HetRec2011 movies have been matched

[13] As opposed to contextualized and/or personalized.
[14] The final results of the UBES and GKG summaries, both using Freebase URIs, can be found in the dataset, cf. Sec. 4.3.

to Freebase identifiers (cf. [9] for more information). Based on the rating data provided by HetRec2011, the 20 nearest neighbors for each of the 60 selected movies were derived with the help of the Apache Mahout[15] library. It has to be noted that the actual numerical ratings were not used due to utilization of the log-likelihood similarity score [5]. This similarity measure only uses binary information (i. e., rated and not rated). With two SPARQL queries per movie, the number of shared features was estimated once in combination with the neighbors and once considering the whole dataset. These numbers enable to apply the TF-IDF-related weighting for each property as it is described in [9]. Finally, the output has been filtered with the white list described in Sec. 3.1 in order to fit with the properties of the game and GKG.

Google's Knowledge Graph (GKG) summaries

The 60 movie summaries by Google have been processed in a semi-automatic way to fit with the Freebase URIs. The first step was to retrieve the summaries of all 60 movies and storing the according HTML files. While the Freebase URIs for properties such as "Director" had to be entered manually, most objects could be linked to Freebase automatically. For this, we made use of the GKG-Freebase link[16]. The ranking of the five main facts is to be interpreted in a top-down order while Google's ordering of 'Cast' members follows a left to right orientation.

4 Results

At present, our quiz has been played 690 times by 217 players, while some players have played more frequently and the majority of 135 players has played only once. All 2,829 triples have been played at least once, 2,314 triples at least three times. In total 8,308 questions have been replied of which 4,716 have been answered correctly. The current results have to be regarded with care, since the absence of multiple opinions about a portion of the facts increases the probability for outliers. The random summaries were generated in accordance to the white list (cf. Sec. 3.1). In order to gain real randomness, we averaged the scores of 100 randomly generated summaries.

The ratio of correctly answered questions varies depending on the property that has been used in the question. As shown in table 1, to determine a movie according to its *prequel*, *film series*, or *sequel* is rather obvious, whereas a *film festival* or *film casting director* does not give a clear idea of the movie in question.

4.1 Evaluation of Property Ranking

To evaluate the ranking of properties for a single movie, we have determined the ranking of properties according to the *correct answer ratio*. The GKG movie representation lists general facts in an ordered manner, whereas the cast of the

[15] Apache Mahout – `http://mahout.apache.org/`

[16] `http://lists.w3.org/Archives/Public/semantic-web/2012Jun/0028.html`

Table 1: Overall Relevance Ranking for Movie Properties

Rank	Property	Correct	Rank	Property	Correct
1	prequel	95.39%	14	production company	56.10%
2	film series	95.16%	15	runtime	54.52%
3	sequel	85.33%	16	music	54.11%
4	parodied	76.47%	17	award	53.41%
5	adapted original	74.32%	18	actor	52.86%
6	subject	73.91%	19	story writer	51.18%
7	genre	65.14%	20	editor	50.00%
8	initial release date	65.14%	21	event	50.00%
9	director	63.51%	22	cinematographer	44.20%
10	rating	61.61%	23	budget	42.78%
11	writer	61.61%	24	film festival	42.27%
12	featured song	60.00%	25	film casting director	41.32%
13	featured filming location	60.00%			

movie is displayed separately. Accordingly, only the remaining 24 properties are used for this evaluation. Properties that do not occur in the systems' results are jointly put in the bottom position. For benchmarking the ordering of both summaries, Kendall rank correlation coefficient is applied. For each movie τ is determined over the set of its properties. Table 2 shows the average, minimum, and maximum findings of τ. It can be seen, that both systems as well as random

Table 2: Performance for Movie Property Ranking for Selected Movies

	τ_{avg}	τ_{min}	τ_{max}
UBES	0.045	-0.505 (The Sixth Sense)	0.477 (Reservoir Dogs)
GKG	0.027	-0.417 (The Big Lebowski)	0.480 (Reservoir Dogs)
Random	0.031	-0.094 (American Beauty)	0.276 (Monsters Inc)

perform equal in average. In each system, for about half of the movies the correlation is negative which means that the orderings are partly reverse compared ordering in the derived dataset. In general, none of the two systems' rankings differs significantly from a random ranking. This might be due to the sparsity of the dataset where most of the facts have been played only three times or less. Another negative influence might come from the fact that we aggregate on objects as we rank properties only and do not consider full property-value pairs.

4.2 Evaluation of Feature Ranking

For this evaluation the relevance ranking of the movie cast is compared to the user generated ground truth. Table 3 presents the average, minimum, and maximum findings of τ for the ranking of actors for a distinct movie. The results for

Table 3: Performance for Actor Ranking for Selected Movies

	τ_{avg}	τ_{min}	τ_{max}
UBES	0.121	-0.405 (The Princess Bride)	0.602 (Indiana Jones and the last Crusade)
GKG	0.124	-0.479 (The Princess Bride)	0.744 (The Matrix)
Random	0.013	-0.069 (Fargo)	0.094 (Good Will Hunting)

the actor ranking are fairly equal for both systems in the average case. The average τ value differs from random scores. We have estimated that the difference to the random ranking is significant ($p < 0.05$) for both systems. This result provides an indication that the relative importance of property-value pairs can be captured by the statistics established through the game. It has to be mentioned, that - in some cases - the UBES heuristic provides none or very few proposals due to the required *'Cast'* overlap to neighboring movies.

4.3 Published Dataset

By publishing the data collected within the game[17], we encourage other researchers to apply this information for their purposes. The dataset consists of two main parts: first the aggregated statistics, which comprises the selected RDF triples and the respective players' performance. And second an anonymized log about the completed games that allows replay of user sessions with complete questions and results. Updates of these files will be published on a regular basis.

5 Conclusion and Future Work

In this paper a crowd sourcing approach implemented as a game with a purpose is demonstrated to gather relevance information about facts within a knowledge base and to establish ground truth data for evaluating summarization. We found indications that such a dataset can fulfill this purpose. However, the established dataset in its current state is too sparse to make valid assumptions about the importance of single facts.

Future development of the *WhoKnows?Movies!* game will also include images to help players to identify persons related to a movie, or other composed information artifacts. We also consider scoring properties that were listed in combination with an incorrect object while the user did not vote for this answer possibility. This is due to the fact that the user probably could exclude this possibility as he knew the correct object(s). Further research directions are increasing the number of movies and exploiting further domains. As for the latter, we consider the domains of books, music, places, and people. In principle, any domain where general knowledge is widely spread can be targeted with the game.

[17] The dataset is available at `http://yovisto.com/labs/iswc2012/`

Acknowledgements

The authors would like to thank Ontotext AD for providing OWLIM-SE 5.0. This research was partly funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 257790 (RENDER project).

References

1. Enrique Amigó, Julio Gonzalo, Anselmo Peñas, and Felisa Verdejo. Qarla: a framework for the evaluation of text summarization systems. In *Proc. of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 280–289, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.
2. Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2nd ws. on information heterogeneity and fusion in recommender systems (hetrec 2011). In *Proc. of 5th ACM Conf. on Recommender systems*, RecSys 2011, New York, NY, USA, 2011. ACM.
3. Gong Cheng, Thanh Tran, and Yuzhong Qu. RELIN: relatedness and informativeness-based centrality for entity summarization. In *Proc. of the 10th int. conf. on The Semantic Web - Vol. Part I*, ISWC'11, pages 114–129, Berlin, Heidelberg, 2011. Springer-Verlag.
4. comScore. comscore reports global search market growth of 46 percent in 2009. `http://www.comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009`, 2010.
5. Ted Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, 1993.
6. H. P. Luhn. The automatic creation of literature abstracts. *IBM J. Res. Dev.*, 2(2):159–165, April 1958.
7. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. of the 10th Int. Conf. on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.
8. Amit Singhal. Introducing the knowledge graph: things, not strings. `http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html`, 2012.
9. Andreas Thalhammer, Ioan Toma, Antonio J. Roa-Valverde, and Dieter Fensel. Leveraging usage data for linked data movie entity summarization. In *Proc. of the 2nd Int. Ws. on Usage Analysis and the Web of Data (USEWOD2012) co-located with WWW 2012, Lyon, France, 2012*, volume abs/1204.2718, 2012.
10. Jörg Waitelonis, Nadine Ludwig, Magnus Knuth, and Harald Sack. WhoKnows? – evaluating linked data heuristics with a quiz that cleans up DBpedia. *Int. Journal of Interactive Technology and Smart Education (ITSE)*, 8(3):236–248, 2011.
11. Jörg Waitelonis and Harald Sack. Towards exploratory video search using linked data. *Multimedia Tools and Applications*, 59:645–672, 2012. 10.1007/s11042-011-0733-1.