

LinkSUM: Using Link Analysis to Summarize Entity Data

Andreas Thalhammer¹, Nelia Lasierra², and Achim Rettinger¹

¹ AIFB, Karlsruhe Institute of Technology, Germany
{andreas.thalhammer, achim.rettinger@kit.edu}

² University for Health Sciences, Medical Informatics and Technology, Austria
nelia.lasierra@umit.at

Abstract. The amount of structured data published on the Web is constantly growing. A significant part of this data is published in accordance to the Linked Data principles. The explicit graph structure enables machines and humans to retrieve descriptions of entities and discover information about relations to other entities. In many cases, descriptions of single entities include thousands of statements and for human users it becomes difficult to comprehend the data unless a selection of the most relevant facts is provided.

In this paper we introduce LinkSUM, a lightweight link-based approach for the relevance-oriented summarization of knowledge graph entities. LinkSUM optimizes the combination of the PageRank algorithm with an adaption of the Backlink method together with new approaches for predicate selection. Both, quantitative and qualitative evaluations have been conducted to study the performance of the method in comparison to an existing entity summarization approach. The results show a significant improvement over the state of the art and lead us to conclude that prioritizing the selection of related resources leads to better summaries.

Keywords: entity summarization, linked data, knowledge graph, information filtering

1 Introduction

A significant part of search engine result pages (SERPs) is nowadays dedicated to knowledge graph panels about entities (e.g., Figure 1). In that context, a large amount of information about searched entities is often readily available to be presented to the user in a structured way. In its complete form, data about a single entity may involve thousands of statements. This is an overloading amount for humans. Therefore, fact-based information is often filtered and presented with a pre-defined set of predicates, such as “*name, age, and date of birth*” in the case of persons. Such a listing is usually associated with fixed patterns and static type-based orderings. However, as each entity is special in its own way it would be more appropriate to select relevant facts with respect to its individual particularities. In the movie domain, for example, some movies are



Fig. 1: Parts of a Google Knowledge Graph summary of “Pulp Fiction”.

heavily influenced by their main actor(s) (e. g., in the case of “Terminator”) while others are genuine masterpieces by their directors (e. g., in the case of “Pulp Fiction”). It is the goal of entity summarization to distill such individual particularities and present them in a ranked fashion.

In the last five years, the field of entity summarization has gained particular attention by both, industry [1,13] and research [2,4,7,14,16,17,18,19]. On the one hand, the commercial approaches are very specific to their individual settings and rely on large amounts of background information. From their interfaces it is also indistinguishable, how much of the approach is automatic and which parts are manually generated or revised. As such, these approaches can neither be generally applied nor reproduced. On the other hand, the approaches from the scientific field are more generic and generally applicable. Among those, we distinguish between *diversity-centered* summaries [7,14] and *relevance-oriented* summaries [4,17,19]:

Diversity-Centered Summaries focus more on presenting a diverse selection of predicates (i. e. the type of relation). Repetitive lists of the same type of relation (e. g., “*starring* Uma Thurman; *starring* John Travolta; *starring*...”) are avoided in this setting. Instead, diversification of the relations aims at providing a more complete overview of an entity.

Relevance-Oriented Summaries are more focused on the values (i. e. the connected resources). The importance of the connected resource and the relevance for the target entity is prioritized. In this setting, a complete summary could involve only one type of relation, if the respective resources are deemed more important than others with different relations.

Both methods present summaries of entities in a top- k manner, i. e. the k most diverse or relevant facts.

In this paper we present LinkSUM, a new method for entity summarization that follows a relevance-oriented approach to produce generic summaries to be displayed in a SERP. LinkSUM goes beyond the state of the art by addressing the following observed limitations of previously developed methods: lack of general applicability (commercial approaches) and the inclusion of redundant information in a summary (commercial and research approaches).

To address these challenges, LinkSUM combines and optimizes techniques for resource selection with approaches for predicate selection in order to provide a generic method for entity summarization. Like other research and commercial approaches [1,4,7,14,19], LinkSUM is focused on global relevance measures

and does not rely on personal or contextual factors like individual interests or temporal trends. Instead, it serves as a foundation which can be extended by such approaches. To study the performance of LinkSUM we compare it with FACES, a recent approach on entity summarization [7] that has been shown to perform better than [4] and [17].

The contribution of this paper is twofold:

1. We present LinkSUM, a lightweight link-based approach for relevance-oriented entity summarization. We investigate on different configuration parameters and evaluate them with respect to their effectiveness.
2. In a quantitative and qualitative evaluation setting we show that prioritizing the selection of the related resources (rather than focusing on relation selection) and omitting redundancies within the set of related resources leads to better summaries.

The remainder of the paper is organized as follows: Section 2 introduces the key components of our approach. Section 3 presents first the experimental setup and afterwards the results of the configuration of the approach. In Section 4 we compare the approach to a diversity-centered summarization approach in a quantitative as well as qualitative evaluation setting. Section 5 discusses the obtained results and Section 6 provides an overview of related work. Section 7 presents our conclusions and Section 8 addresses open topics that will be part of our future work.

2 Proposed Approach

The proposed entity summarization method comprises two main stages:

Resource Selection The goal of this stage is to create a ranked list of resources that are semantically connected to the target entity. The output of this step is a set of triples, where the semantic relation is not fixed, e. g.

Pulp Fiction – ?relation → *Quentin Tarantino*. One requirement for a resource to be included in the list of relevant entities is at least one existing semantic relation to the target entity.

Relation Selection This stage deals with the selection of a semantic relation that connects the resource with the target entity. This step is necessary if more than one relation exists, e. g.

Pulp Fiction – *starring* → *Quentin Tarantino*, and
Pulp Fiction – *director* → *Quentin Tarantino*.

In the entity summarization setting the list of relevant resources is cut-off at k after resource selection. We refer to such summaries as top- k summaries. In the following subsections we will explain each of the two parts. We will refer to the target entity as e (i. e. the entity that needs to be summarized).

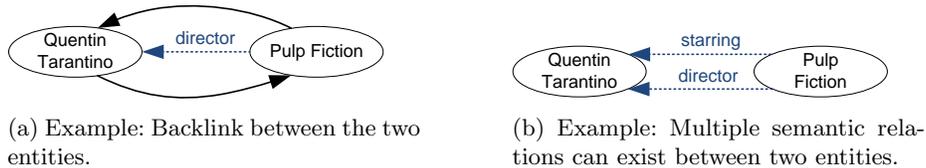


Fig. 2: Web links (black, solid) and semantic relations (blue, dashed) between “Quentin Tarantino” and “Pulp Fiction”.

2.1 Resource Selection

For the resource selection, we combine two link-measures: one that accounts for the importance of the connected resource (PageRank [3]) and one that accounts for the strength of the connection (Backlink [20]). We consider links between entities as a mean for identifying and ranking relevant resources. The presented method covers scenarios, in which semantic relations are present in addition to textual descriptions that contain Web links to other resources.

Important Related Resources As a first step, we run the PageRank algorithm [3] on the set of all resources R with their individual directed links $link(r_1, r_2)$ with $r_1, r_2 \in R$ and their individual count of out-going links: $c(r) = |\{r_1 | link(r, r_1); r_1 \in R\}|$.

$$pr(r_0) = (1 - d) + d \cdot \sum_{r_n \in \{r | link(r, r_0); r \in R\}} pr(r_n) / c(r_n) \quad (1)$$

The variable d is called “damping factor”. Generally, it accounts for the probability of a jump in the random walk model of PageRank. Like in [3], we set the damping factor to 0.85 in all our experiments. The PageRank algorithm applies the above-given formula incrementally. The number of iterations depends on the general size of the dataset as well as on the density of links. After executing the algorithm, each resource r has its own PageRank score $pr(r)$. The set of resources that have a semantic connection to e is defined as $res(e) \subseteq R$. As a matter of fact, every resource $r \in res(e)$ can be ranked in accordance to its individual PageRank. A basic popularity-based top- k summary of e can be produced with that information [17].

Strongly Connected Resources PageRank focuses on the general importance of related resources. It does not provide an indication about how the two resources are important for each other. This part is addressed by the Backlink method that was first described in [20]. In this work, the authors analyze a variety of set-based heuristics for identifying related resources in order to feature exploratory search with Linked Data. The analyzed Backlink method performs best in terms of F-measure when the results are compared to their reference

dataset. In [20] the method is introduced as follows:

$$bl(e) = \{r | link(r, e) \wedge link(e, r), r \in R\} \quad (2)$$

For entity summarization, we adapt the Backlink method in order to ensure that a semantic relation exists between e and every r . The adapted formula is as follows:

$$bl(e) = \{r | link(r, e) \wedge link(e, r) \wedge r \in res(e), r \in R\} \quad (3)$$

Figure 2a shows the Backlink method and the additional requirement for a semantic relation between two resources. Backlink can not be used directly for entity summarization as it returns an unranked set of related entities and the size of this set depends on the target entity.

Combined Scores for Resource Selection In this work, an optimized combination of PageRank with Backlink is proposed. This enables us to select relevant resources with a tight connection to e . For this, we normalize the PageRank score of each entity by the maximum and linearly combine the score with the indicator function of the set $bl(e)$. With $r \in res(e)$:

$$score(e, r) = \alpha \cdot \frac{pr(r)}{\max\{pr(a) : a \in res(e)\}} + (1 - \alpha) \cdot \mathbf{1}_{bl(e)}(r) \quad (4)$$

For a top- k summary we rank resources $r \in res(e)$ in accordance to the defined score and cut off at k . We define a top- k list of connected resources with the function $top_k(res(e))$. The α parameter is flexible and lies in the interval $0.5 \leq \alpha \leq 1$. With $\alpha = 0.5$, the top positions of a summary of e first involve all resources contained in the Backlink set $r \in bl(e)$ in the order of their PageRank scores. This listing is followed by the resources that are not in the Backlink set $r \notin bl(e)$ but still semantically connected $r \in res(e)$ in the order of their PageRank scores. This is also the case if α is chosen in the interval $0 < \alpha \leq 0.5$. With $\alpha = 1.0$, all connected resources $r \in res(e)$ are ordered in accordance to their PageRank scores. In this case, the Backlink set does not influence the results. In Section 3 we present different configurations of LinkSUM with respect to α .

2.2 Relation Selection

In a knowledge graph, two resources can be linked through multiple semantic connections. We provide an example in Figure 2b which demonstrates that the entities “Pulp Fiction” and “Quentin Tarantino” are connected in multiple ways. As a matter of fact, it is very common that multiple relations between entities exist. However, in many cases, one relation is more relevant than others. In our approach, the relation selection task identifies the most prominent connection for presentation in order to avoid redundancies among the connected resources in the top- k set.

In order to choose an optimal relation selection method for LinkSUM, the following factors were defined:

Frequency (FRQ) Ranks the candidate relations in accordance to how often a specific relation is used overall in the complete dataset. The relation that is used the most is selected as the most promising candidate.

Exclusivity (EXC) For both entities of a relation, the relation might not be exclusive. For example a movie has commonly more than one starring actor while also an actor is usually starring in more than one movie (N:M). This measure considers the exclusivity of a relation in context to e and $r \in res(e)$ respectively. For both resources, e and r , we add up the number of times the relation is used with each (N+M). We use the inverse of this number $1/(N + M)$, in order to get the exclusivity score (the more exclusive, the better). The upper bound of EXC is 0.5 (for a 1:1 relation).

Description (DSC) Relations are represented by RDF predicates. Those predicates are commonly described with domains, ranges, and labels in different languages. The sum $|labels| + |ranges| + |domains|$ forms a basic method for estimating the quality of the description of the predicate. The relation with the highest quality is chosen.

For each related resource in $r \in top_k(res(e))$, combinations of the above-presented relation selection mechanisms identify the most relevant connection to e .

3 Configuration

As reported in [7], the FACES system (to which we compare) was tuned to its best performance by setting the cut-off level of the cluster hierarchies to 3. Also LinkSUM can be configured with respect to various parameters. First, the α -value for resource selection is flexible in the range of 0.5 to 1 (see Section 2.1). Second, the relation selection method can be adjusted or replaced in order to fit one or another scenario (see Section 2.2). For finding the best configuration, we considered the following configurations:

α -value We tested different settings for α in the range of 0.5 to 1 with 0.1 steps.

Relation Selection We tested different relation selection mechanisms. We considered only combinations based on frequency as it has been proven as a robust popularity measure in [14]. The following setups were considered as promising candidates:

- FRQ – relations are selected by their frequency in the dataset.
- FRQ*EXC – relations are chosen by the product of frequency and exclusivity.
- FRQ*DSC – relations are selected by the product of frequency and description.
- FRQ*EXC*DSC – relations are chosen by the product of frequency, exclusivity, and description.

As a reference dataset, we use the same as the FACES approach [7].³ The dataset provided in [2] would also serve as reference for evaluation. Unfortunately, we could not obtain summaries of the FACES system for the entities covered by [2].

The dataset provided by FACES involves DBpedia (version 3.9) and features outgoing connections only [7]. Without loss of generality, we also configured LinkSUM to consider outgoing connections only. We also apply LinkSUM on DBpedia version 3.9. We computed the PageRank [3] scores for each DBpedia entity. As a basis for this, we used DBpedia’s Wikipedia Pagelinks dataset in English language. This dataset contains triples of the form “Wikipedia page A links to Wikipedia page B”. We only use these Web links, i.e. do not make use of semantic links (e.g., `dbpedia-owl:birthPlace`) for the computation of PageRank. The computed PageRank scores are made available online [15] in Turtle RDF format using the vRank vocabulary [11]. For the Backlink method, we also use the Wikipedia Pagelinks dataset.

3.1 Configuration Setup

Our experimental setup involves a reference dataset as well as measures for computing the agreement and similarity. We use a similar evaluation setup as the FACES approach [7] as we directly compare LinkSUM with the FACES system (see Section 4).

Reference Dataset The dataset includes 50 DBpedia (version 3.9) entities. The dataset contains at least seven top-5 and seven top-10 reference summaries per entity that were created by 15 experts of the Semantic Web field [7]. For each entity, these references describe outgoing connections to other resources. The average number of these relations is 44. In addition, several relations, such as `dcterms:subject` and Wikipedia related links, were filtered out for creating the reference dataset as they do not contain sufficient semantic information [7].

Measures For computing the agreement and for comparing the produced summaries with the reference dataset, we use the same similarity measures as in [4] and [7]:

$$Agreement(e) = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n |Sum_i^E(e) \cap Sum_j^E(e)| \quad (5)$$

$$Quality(Sum(e)) = \frac{1}{n} \sum_{i=1}^n |Sum(e) \cap Sum_i^E(e)| \quad (6)$$

With n being the number of experts. The expert summaries are denoted as $Sum_i^E(e)$. The agreement measure estimates the agreement of the experts

³ FACES reference dataset – <http://wiki.knoesis.org/index.php/FACES>

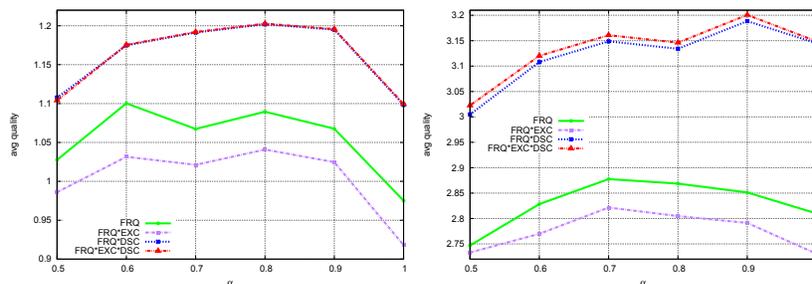


Fig. 3: LinkSUM (SPO) average *Quality* scores with different settings for α and different relations selection approaches for top-5 (left) and top-10 (right) summaries.

about a top- k summary of the entity e . The *Quality* measure estimates the overlap of the produced summary $Sum(e)$ with all expert summaries. Both values are computed for all entities in the reference dataset and afterwards averaged.⁴ The upper and lower bounds for both measures are $0 \leq Agreement(e) \leq k$ and $0 \leq Quality(Sum(e)) \leq k$ in the top- k setting. When we reproduced the setting of [7], we found that our results did not match the provided values: the *Quality* values of FACES were lower than the provided ones. In order to reproduce the reported values for the FACES system in [7], we found out that only the last part of the URI was used for matching automatically generated summaries with expert summaries for all tested systems. Unfortunately, this setting matches DBpedia predicates with different namespaces (i. e. `dbpprop` and `dbpedia-owl`) in an arbitrary way. As an example, on the one hand, `dbpprop:party` and `dbpedia-owl:party` are matched while, on the other hand, `dbpprop:placeOfBirth` and `dbpedia-owl:birthPlace` remain unmatched because the last parts of the URI are syntactically not the same. As a consequence, we decided not to adopt this basic ontology alignment approach and applied two measures instead:

- Subject–Object (SO): This measure treats a summary as a set of tuples containing only subjects and objects while ignoring the predicate. The full URIs of the subject and the object are used respectively. As a matter of fact, the relation selection method has no impact on this measure.
- Subject–Predicate–Object (SPO): This measure treats summaries as sets of triples. For representing a triple we use the full URI of each, subject, predicate and object. This measure also estimates the performance of the relation selection approach.

⁴ k is fixed to the same value for all summaries, expert and automatically generated ones, before applying the measures.

Table 1: Agreement among the experts.

	SO	SPO
top-5	2.14	1.64
top-10	5.14	3.92

3.2 Configuration Results

In [7], the reported agreement among the experts is 1.92 for top-5 and 4.64 for top-10 respectively. Those values were computed with the aforementioned basic ontology alignment approach. We recomputed the values for SO and SPO respectively. The results are displayed in Table 1. The agreement among the experts is not particularly high. According to [7], this can be explained by the high number of facts that were presented to the experts for each entity (in average 44 facts per entity). Although - technically - the average agreement is not an upper bound for the performances of the tested systems, the values can serve as reference points.

In the SO setting, the best achieved scores of LinkSUM are 1.89 (top-5, $\alpha = 0.8$) and 4.82 (top-10, $\alpha = 0.9$) respectively. The results of the SPO settings are shown in Figure 3. The FRQ measure provides a good baseline for both, top-5 and top-10. While the combination of FRQ with DSC improves the *Quality* in both settings, the combination with EXC damps the impact of FRQ. In the top-10 setting, the combination of the three measures (FRQ*EXC*DSC) provides best values. In the top-5 settings, FRQ*DSC and FRQ*EXC*DSC provide equally good results. In general, the values for α are best at 0.8 for top-5 and 0.9 for top-10. The impact of the Backlink method on the rankings ($\alpha < 1.0$) in comparison to PageRank-only ($\alpha = 1.0$) is evident. In addition it is noticeable that strictly prioritizing all results of the Backlink method (ranked in accordance to their respective PageRank scores) does also not yield best results ($\alpha = 0.5$). The full blend between importance and strong connectivity produces the best outcomes.

Summarizing, the following configurations performed best for top-5 and top-10 summaries respectively:

config-1 (top-5): $\alpha = 0.8$, FRQ*EXC*DSC

config-2 (top-10): $\alpha = 0.9$, FRQ*EXC*DSC

4 Evaluation

In our evaluation setting, we compare LinkSUM with the FACES entity summarization system [7]. FACES focuses on the diversification of the relation types (i. e. no semantically similar predicates should occur in the result summary). The system has two stages: partitioning the feature set and ranking the features. The main idea is to partition the semantic links of an entity into semantically diverse clusters of predicates. For resource selection, the approach uses a tf-idf-related popularity measure for the object. In contrast, in our approach we follow

the objective to identify the most relevant object first and then select the predicate. In their evaluation, the authors demonstrate that their system provides better results than [4] and [17]. For 50 DBpedia entities, the authors published the results of FACES for top-5 and top-10 summaries (along with the reference dataset described in Section 3.1).⁵ The used DBpedia version is 3.9.

We compare LinkSUM and FACES in two evaluation settings, a quantitative and a qualitative one. In the following we will first describe the experimental setup and the obtained results afterwards.

4.1 Evaluation Setup

Quantitative Analysis For evaluating the two methods quantitatively, we chose the same setup as described in Section 3.1, i. e. the same reference dataset and the same evaluation measures that were used for the evaluation of the FACES system [7]. For comparison, we use the average *Quality* of each method. In addition, in order to prevent influence of strong outliers, we use the *Quality* value of each of the 50 entities per system for computing significance. As a significance test, we use the Wilcoxon Signed-Rank Test with two tails as recommended in [5]. We compare the best configurations of LinkSUM for top-5 and top-10 respectively (see Section 3.2) with the published results of FACES.

Qualitative Analysis For the qualitative evaluation we sent a call for participation to more than 60 people and asked them to compare summaries of different entities. In this setup, we evaluated the top-10 setting with LinkSUM@config-2 (which turned out to perform best for the top-10 setting in the configuration, see Section 3.2). We chose a set of ten entities out of the 50 provided summaries of FACES with respect to their types. The types of the selected entities involve the following classes: person, country, football club, TV series, movie, and company. The selection between the entities of a specific type was random.

For each entity, we displayed the summaries of the two systems next to each other (see Figure 4) without giving indications about which system produced the summaries. The summaries produced by LinkSUM were displayed on the left side in 50% of the cases with random choice. Below each summary, we provided a radio button for the users to choose their preferred summary. Every user had one vote either for LinkSUM or FACES. We used two 5-point Likert scale questions in order to enable participants to provide information about their previous knowledge about the entity and the confidence with their choice:

- “*I know a lot about this entity*” – [Strongly agree; Agree; Neither agree, nor disagree; Disagree; Strongly disagree]
- “*I am sure that I prefer the chosen summary over the other*” – [Very confident; Confident; Neutral; Not very confident; Not at all confident]

⁵ FACES summaries – <http://wiki.knoesis.org/index.php/FACES>

The Cosby Show

I know a lot about this entity:

- Strongly agree Agree Neither agree, nor disagree Disagree Strongly disagree

language	English language	network	NBC
network	NBC	language	English language
genre	Sitcom	opening theme	Bobby McFerrin
camera	Multiple-camera setup	composer	Bill Cosby
starring	Bill Cosby	executive producer	Tom Werner
opening theme	Bobby McFerrin	distributor	Paramount Domestic Television
related	A Different World	genre	Sitcom
executive producer	Tom Werner	creator	Bill Cosby
starring	Malcolm-Jamal Warner	company	Viacom Productions
starring	Lisa Bonet	format	480i

>> <<

I am sure that I prefer the chosen summary over the other:

- Very confident Confident Neutral Not very confident Not at all confident

Please provide reasons why you prefer the chosen summary over the other (optional):

e.g.
- reason 1
- reason 2

Fig. 4: Excerpt of the interface for qualitative evaluation for the entity “The Cosby Show”. The users could choose whether they prefer the summary of LinkSUM (left) or FACES (right) in a SERP setting.

Besides we provided an optional field where comments about their choice could be given. We included the following introductory text in order to instruct the users on how to proceed with the evaluation:

“You have been searching on a Web search engine for an entity. The search engine result page (SERP) is displayed with a picture of the entity, a short textual description, and a box with facts about the entity. For the following ten entities, it is your task to decide which fact box you would like to see in a SERP.”

In addition, we asked the participants to assume that all displayed data is correct. This was to avoid influence of data quality on the results. Finally, for statistical classification, we requested the participants to provide the following information: gender, age, whether or not the participants had a background in computer science, and the time taken for evaluation.

4.2 Evaluation Results

In the following, we present the outcomes of both evaluation settings.

Quantitative Analysis In Table 2, we present the overall *Quality* results of the quantitative evaluation. In average, both configurations of LinkSUM achieve better results than FACES in the described settings (top-5/top-10, SO/SPO). LinkSUM@config-2 performs significantly better than FACES in all settings ($p <$

Table 2: Overall *Quality* results of the quantitative evaluation and their respective standard deviation (SD). Best results are bold. † compared to the best, difference is significant ($p < 0.05$); ‡ compared with each of the other two settings, difference is significant ($p < 0.05$).

	SO (top-5)	SPO (top-5)	SO (top-10)	SPO (top-10)
LinkSUM@config-1	1.89 (SD 0.55)	1.20 (SD 0.57)	4.78 (SD 1.05)	3.15 (SD 0.89)
LinkSUM@config-2	1.84 (SD 0.60)	1.20 (SD 0.60)	4.82 (SD 1.06)	3.20 (SD 0.87)
FACES	1.66 [‡] (SD 0.57)	0.93 [‡] (SD 0.54)	4.33 [‡] (SD 1.01)	2.92 [†] (SD 0.94)

0.05). LinkSUM@config-1 is significantly ($p < 0.05$) better than FACES in three of four settings while the level of significance is not fully reached at SPO, top-10.

Qualitative Analysis From the invited people, a total of 20 participated in the qualitative analysis. 75% of the participants were between 25 and 35, and 25% were between 35 and 45 years old. 75% were male and 25% were female. 95% of the participants had a computer science background. The average time taken for the evaluation was 11 minutes and 27 seconds. In total, 13 participants used the option to comment about their choice. With respect to these characteristics, we did not find any significant difference within the distribution of the votes. The distribution of the votes is visualized in Figure 5. 73% of all votes were given to LinkSUM, 27% of the votes were received by FACES. Out of ten entities, LinkSUM system was clearly chosen with more than 15 votes in the case of five entities. For another 2 entities, the LinkSUM system was chosen with votes in the interval 13 to 14. The votes for the remaining three entities were distributed in the interval of 9 to 11 for both systems. Both systems each received in total ten low-confidence votes (“*Not very confident*” or “*Not at all confident*”). This means that 10 out of 146 votes in the case of LinkSUM, and 10 out of 54 votes in the case of FACES were low-confidence votes. With respect to the total number of votes for each system, this means a disproportionate low number of low-confidence votes for LinkSUM. The amount of knowledge of the participants did not influence the preference for either system: the values for high or low knowledge were both in line with the total distribution of the votes.

Another interesting part of the results of the evaluation are the comments of the participants. We group the comments into two categories depending on hints about the decision-making process of the participants. In many cases, the participants gave reason why they selected a summary and/or why they rejected the other. The most-provided reasons for selection/rejection were as follows:

Selection the presented resources are relevant for the entity (e. g., “I like to see Turing machine mentioned for Alan Turing”).

Rejection redundancy (e. g., “The same thing twice once with prize and once with award”), the presented resources do not characterize the entity (e. g., “I do not care about technical aspects such as format”).

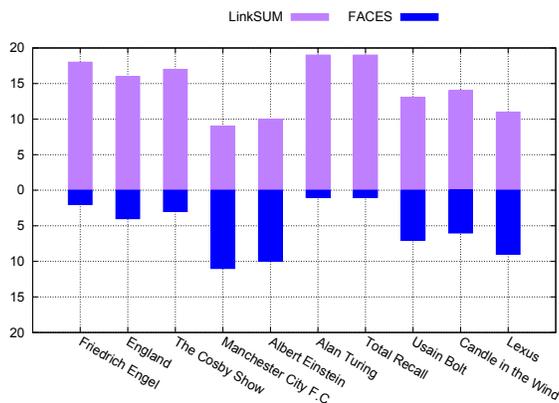


Fig. 5: Results of the qualitative evaluation. The x-axis denotes the respective entities and the y-axis accounts for the number of user votes per system.

5 Discussion

To select the most relevant facts that characterize an entity is, in many cases, a subjective task. Thus, to produce a generic summary not tailored to any specific background or context the user might have is a challenging task that involves the identification of facts that are deemed important by the majority of the users. In order to address this challenge, the LinkSUM method combines and optimizes methods that enable to select relevant facts about entities and at the same time reduce the amount of redundant information. In our experiments and evaluation we assessed and analyzed the efficiency of the mentioned aspects of the LinkSUM method. In a quantitative as well as qualitative setting we compared LinkSUM to the FACES system. In both setups, we demonstrated that LinkSUM exhibits significantly better results than FACES. The comments of the participants of our qualitative experiment suggest that the relevance of the related resources should be of importance and at the same time characterize an entity. We cover this by the combination of PageRank with Backlink. Our experiments with the SO-measure demonstrate that the produced *Quality* values are close to the agreement of the expert summaries (cf. Table 1).

We have tested four different methods for relation selection. The combination of the frequency of the relation, its exclusivity, and the its description has been shown to perform best in the top-10 setting, while in the top-5 setting the exclusivity score did neither contribute positively, nor negatively in that setup. The introduced measures should be considered as baselines for future evaluation settings in context to the relation selection step.

With regard to the qualitative evaluation, in the cases of the entities “Manchester City F.C.”, “Albert Einstein”, and “Lexus” we could not find any clear majority for either of the two systems. In the case of “Lexus” the set of presented facts has a very high overlap between the systems (with different ordering). In

the case of “Manchester City F.C.” and “Albert Einstein” the choices are subjective as the provided comments suggest: some users liked the listing of players (“Manchester City F.C.”) or children (“Albert Einstein”) while others stated they did not. Contrary to the claims in [7], we could not find evidence that repetitive relations have a negative impact on the quality of the summaries. For example, the entity “The Cosby Show” contains a listing of various actors with the “starring”-relation in the LinkSUM summary while in the output of FACES this information is missing (see Figure 4). This led to 17 vs. three votes for the LinkSUM method. In this case many of the participants provided the “inclusion of the actors” in the LinkSUM method as the main reason for their choice. The FACES system does not filter redundancies on the object level: it happens that the set of relations is diverse while on the object side, a connected resource is re-occurring multiple times (linked through different relations). An example is the entity “Total Recall (1990 film)” where FACES included the following information: *director Jerry Goldsmith; Artist Jerry Goldsmith; music Jerry Goldsmith; music composer Jerry Goldsmith; screenplay David Cronenberg; writer David Cronenberg*. Those and similar repetitions in the summaries of other entities were commented as “redundant” by a high number of participants (in total ten out of 13 participants with comments mentioned redundancy as a problem).

At <http://km.aifb.kit.edu/services/link>, a deployment of LinkSUM is available online. It implements the SUMMA entity summarization API [18].

6 Related Work

To the best of our knowledge, Hogan et al. first mentioned the concept of “summaries of the relevant entities” in [8].

The authors of [4] introduce RELIN, a summarization system that supports quick identification of entities. The approach 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 of the contribution is to combine textual notions of informativeness and relatedness for the ranking of features. As a major effect, the concise presentation of retrieved entities for quick identification by users after search is one of the scenarios that RELIN supports. In [7], the system is shown to perform significantly worse than FACES.

Google “Knowledge Graph” [13] is an example for an entity search system. The main idea is to enrich search results with summarized information about named entities. While the details of the approach are not public, Amit Singhal, the author of [13], outlines that for summarization, the system goes back to user data in order to “... study in aggregate what they’ve been asking Google about each item”. This indicates, that Google uses additional data sources for the summaries, i. e. the queries of the users. In addition, this also provides reason to assume that the analysis focuses on informal and partial statements of the subject+predicate or subject+object kind. Our approach is similar to this methodology and follows the pattern of identifying important objects first and then select a predicate.

TripleRank by Franz et al. introduces a tensor-based approach for ranking RDF triples [6]. The approach uses the PARAFAC tensor decomposition method for deriving authority and hub scores as well as information about the importance of the link type. In contrast, in our contribution we separate the steps of deriving importance of the resource and the importance of the link as we put additional focus on the context that the target entity brings (while TripleRank addresses a more general ranking of triples). However, our general PageRank importance scores can be easily augmented or replaced by the scores produced by the TripleRank method.

The authors of [14] discuss the notion of diversity for graphical entity summarization. Two algorithms are introduced, of which one is diversity-oblivious (called PRECIS) and the other is diversity-aware (called DIVERSUM). The evaluation of the algorithms is shaped towards the movie domain and involved expert-based assessments as well as crowd-sourced experiments. The results suggest that the DIVERSUM algorithm was favored over the PRECIS algorithm. A drawback of the method is that both algorithms treat the predicate-value pairs on a per-predicate basis without measures on the object.

Also with respect to diversity, Schäfer et al. detect anomalies about entities in accordance to their different types [12]. At the current state, the system needs also the specific type as an input. However, if the main type of an entity is detected reliably, the method can be regarded as an entity summarization system that points out hidden or interesting facts.

Blanco et al. introduce Yahoo!'s Spark system [1], an entity recommendation system that suggests related entities based on a learning approach employing gradient boosted decision trees. The utilized features range from co-occurrence information over popularity features (such as the click frequency) to graph-theoretic features (such as PageRank). The system focuses on related entity recommendation in the domains of locations, movies, people, sports, and TV shows. The types of entities as well as the type of their relation play an important role in the recommendation process. Connecting predicates are not considered by Spark. The system is currently applied in the Yahoo! search system.

In another contribution [19], Thalhammer et al. exemplify a summarization approach for movie entities that utilizes rating data from the MovieLens dataset. For this, an item neighborhood is established through an item-based collaborative filtering approach. The approach is based on the idea that the semantic background that connects a movie with its neighbors can be found and extracted by making use of structured data. Similar to [4], the authors treat the object and the predicate as predicate-value compounds. The method introduces a tf-idf-based weighting scheme in order to penalize features that occur commonly in the whole dataset.

Waitelonis and Sack explain in their paper [20] how different heuristics can be used for discovering related entities in order to support exploratory search. The tested Backlink heuristic achieves the best results in terms of F-measure. In our contribution, we adopted this method and adapted it in order to fit the scenario of entity summarization. Like all tested heuristics of [20], Backlink provides an

unranked set of related entities that is not directly useable in top- k settings. As a consequence, for our resource selection approach, we combine Backlink with PageRank [3].

In this work, we extended on the state of the art in the field of relevance-oriented entity summarization systems [4,19] and fact ranking in general [6]. Our contribution provides a clear cut between relevance-oriented and diversity-centered systems. We demonstrate that relevance-oriented systems provide a better foundation for displaying summaries in search engine result pages.

7 Conclusions

We presented LinkSUM, a generic relevance-centric summarization method for entity summarization. LinkSUM works with a lightweight two-stage approach in order to produce summaries for entities. In the first step, the method identifies relevant connected resources. In the second step, the system selects the most promising semantic relation for each of the connected resources. We also investigated on the most efficient configuration parameters for LinkSUM.

The results of our quantitative and qualitative evaluation, where we compared LinkSUM to the state-of-the-art system FACES [7], lead us to the following conclusions:

- For SERP scenarios, summarization systems should primarily focus on the relevance and the strength of the connection to the related resources. As a second factor the selection of an appropriate semantic relation is of importance.
- Redundancies in the set of related resources should be avoided (e.g., see Figure 1). Commonly, if two entities are related, there is one relation that is more relevant to be mentioned. Summaries should focus on this relation and then present relations to other interesting resources.

We demonstrated applicability of the LinkSUM method for the DBpedia and Wikipedia datasets and provide results that significantly improve the state of the art. The LinkSUM system is relevant to many other tasks, like e.g.,

Semantic MediaWiki Semantic MediaWiki (SMW) [9] is a popular extension of the MediaWiki software (used by Wikipedia). In SMW, (hyper-) textual information about entities is combined with structured information about them. Using the hyperlinks of the MediaWiki articles in combination with the semantic links of the SMW, LinkSUM can be used to provide structured summaries of entities in SMW.

Microdata/RDFa The number of Web pages that include semantic information about entities is on the rise [10]. In many sites that focus on specific entities, hyperlinks and semantic links are occurring side by side. A prominent example for such co-occurrence is IMDb⁶. Applied in a Web data setting, LinkSUM can use plain hyperlinks in combination with the hidden semantic

⁶ IMDb – <http://www.imdb.com/>

information for providing structured summaries of entities that occur on the Web.

LinkSUM is applicable to both of the above-mentioned scenarios and it remains a technical task to implement prototypes. With respect to research, the DBpedia/Wikipedia setting is the most suitable scenario for evaluation as other researchers can also use the same datasets for providing their own summaries and compare them to LinkSUM (that is available online).

Note that the field of entity summarization is not limited to SERPs. As the availability of structured data is growing, applications for different domains and purposes emerge. Examples include business intelligence, e-learning, health information systems, news pages, data sheets, recipes etc. In fact, this includes all domain-specific cases where it is necessary for users to efficiently comprehend large information resources. In addition, entity summarization systems may adapt to user-context factors such as geo-location, cultural background, or time. As entities are retrieved without a specific information demand (like it is the case in question answering) the full personalization/contextualization of entity summaries remains an open challenge.

The above and further challenges need to be addressed in the emerging field of entity summarization. The LinkSUM method can serve as a generic foundation for such domain and/or user-centric scenarios.

8 Future Work

LinkSUM provides high-quality summaries and improves on the performance of existing solutions in the literature. In order to further improve its performance, to address limitations, and account for new features, we plan investigate on the following open points:

- While in this paper we have presented the evaluation of LinkSUM for the case of generic search in the Web, the performance of the LinkSUM method is planned to be evaluated in specific domain settings (e. g., health information).
- LinkSUM can be combined with a learning-to-rank approach with respect to the α -value and different linear combinations of the predicate selection methods.
- In future versions of LinkSUM, we plan to include literal values - such as strings or dates - as descriptors of the entities. The blending of entity-literal and entity-entity relations into a single summary will receive specific attention.

Acknowledgement. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 611346 and by the German Federal Ministry of Education and Research (BMBF) within the Software Campus project “SumOn” (grant no. 01IS12051).

References

1. R. Blanco, B. B. Cambazoglu, P. Mika, and N. Torzec. Entity Recommendations in Web Search. In *The Semantic Web - ISWC 2013*. Springer, 2013.
2. T. Bobić, J. Waitelonis, and H. Sack. FRanCo - A Ground Truth Corpus for Fact Ranking Evaluation. In *Joint Proceedings of SumPre and HSWI 2015, co-located with the 12th Extended Semantic Web Conference*, volume 1556. CEUR-WS, 2016.
3. S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. In *Proc. of the 7th intl. conf. on World Wide Web 7*. Elsevier, 1998.
4. G. Cheng, T. Tran, and Y. Qu. RELIN: relatedness and informativeness-based centrality for entity summarization. In *The Semantic Web - ISWC 2011*. Springer, 2011.
5. J. Demšar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7:1–30, 2006.
6. T. Franz, A. Schultz, S. Sizov, and S. Staab. TripleRank: Ranking Semantic Web Data by Tensor Decomposition. In *The Semantic Web - ISWC 2009*. Springer, 2009.
7. K. Gunaratna, K. Thirunarayan, and A. P. Sheth. FACES: Diversity-Aware Entity Summarization Using Incremental Hierarchical Conceptual Clustering. In *Proc. of the 29th AAAI Conf. Artificial Intelligence, 2015, Austin, Texas, USA.*, 2015.
8. A. Hogan, A. Harth, J. Umrich, and S. Decker. Towards a Scalable Search and Query Engine for the Web. In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, pages 1301–1302, New York, NY, USA, 2007. ACM.
9. M. Krötzsch, D. Vrandečić, and M. Völkel. Semantic MediaWiki. In *The Semantic Web - ISWC 2006*, volume 4273, pages 935–942. Springer Berlin Heidelberg, 2006.
10. R. Meusel, P. Petrovski, and C. Bizer. The WebDataCommons Microdata, RDFa and Microformat Dataset Series. In *Semantic Web - ISWC 2014*. Springer, 2014.
11. A. Roa-Valverde, A. Thalhammer, I. Toma, and M.-A. Sicilia. Towards a formal model for sharing and reusing ranking computations. In *Proc. of the 6th Int. Workshop on Ranking in Databases in conjunction with VLDB 2012*, 2012.
12. B. Schäfer, P. Ristoski, and H. Paulheim. What is Special about Bethlehem, Pennsylvania? Identifying Unusual Facts about DBpedia Entities. In *Proceedings of the ISWC 2015 Posters & Demonstrations Track*, 2015.
13. A. Singhal. Introducing the knowledge graph: things, not strings. <http://goo.gl/kH1NKq>, 2012.
14. M. Sydow, M. Piłkuła, and R. Schenkel. The notion of diversity in graphical entity summarisation on semantic knowledge graphs. *Journal of Intelligent Information Systems*, pages 1–41, 2013.
15. A. Thalhammer. DBpedia PageRank dataset. Downloaded from http://people.aifb.kit.edu/ath#DBpedia_PageRank, 2016.
16. A. Thalhammer, M. Knuth, and H. Sack. Evaluating Entity Summarization Using a Game-Based Ground Truth. In *The Semantic Web - ISWC 2012*. Springer, 2012.
17. A. Thalhammer and A. Rettinger. Browsing DBpedia Entities with Summaries. In *The Semantic Web: ESWC 2014 Satellite Events*, pages 511–515. Springer, 2014.
18. A. Thalhammer and S. Stadtmüller. SUMMA: A Common API for Linked Data Entity Summaries. In *Engineering the Web in the Big Data Era*. Springer, 2015.
19. A. Thalhammer, I. Toma, A. J. Roa-Valverde, and D. 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)*, 2012.
20. J. Waitelonis and H. Sack. Towards exploratory video search using linked data. *Multimedia Tools and Applications*, 59:645–672, 2012. 10.1007/s11042-011-0733-1.