Test Collection for Evaluating Actionable Knowledge Graphs

Roi Blanco
University of A Coruña, Spain
rblanco@udc.es

Hideo Joho
University of Tsukuba, Tsukuba, Japan
hideo@slis.tsukuba.ac.jp

Adam Jatowt
Kyoto University, Kyoto, Japan
adam@dl.kuis.kyoto-u.ac.jp

Haitao Yu
University of Tsukuba, Tsukuba, Japan
yuhaitao@slis.tsukuba.ac.jp

Abstract

Knowledge graphs (KG) can be used to enrich traditional search results by inserting brief answers to directly respond to users' search needs. As the user needs on search engine diversify, the range of needs answered by KB should also be diversified. However, the resources for developing and evaluating KG generation technologies are still limited. In this paper we discuss the NTCIR-13 Actionable Knowledge Graph (AKG) task and its test collections. The task focuses on finding possible actions related to input entities as well as the relevant properties of such actions. The NTCIR-13 AKG test collections include queries, entities, entity types, set of possible actions for entities, and relevant entity attributes. Finally, we discuss future directions for generating and evaluating actionable KGs.

1 Introduction

Knowledge graphs (KGs) have become an increasingly common and important component in search engine result pages (SERPs). Thanks to knowledge present in the Web, search engines can directly return to users relevant information (alongside web pages), saving user effort in extracting and summarizing data. It is now commonplace for search engines to react to entity-centric user queries using KGs by returning factoid type information about entities (e.g., birthday of a celebrity, restaurant address with a pointer on a map) alongside with related entity items [2] which accompany the traditional blue links and other media nuggets. However, it is well known that users employ search engines not only for acquiring information but also for completing actions and goals [5]. Hence, generating readily actionable output should increase users' satisfaction. Equipped with the collected information on the range of possible actions for a given entity, search engines could display actionable information that corresponds to the most probable underlying search intent behind user queries. Users could then directly act on such output data to more effectively and efficiently complete their desired actions. Furthermore, direct links to services allowing the execution of such actions could be included as a further means for improving the search experience. Given that on average 43% of search queries contain an entity [7], effective solutions for

In: L. Dietz, C. Xiong, E. Meij (eds.): Proceedings of the First Workshop on Knowledge Graphs and Semantics for Text Retrieval and Analysis (KG4IR), Tokyo, Japan, 11-Aug-2017, published at http://ceur-ws.org

supporting entity-centric actions have high potential to facilitate search on the Web. Although there has been considerable research on entity-centric search [6, 7], few proposals investigated the possibility of automatically deriving actions related to entities in search queries for the purpose of search improvement.

In this paper we introduce the concept of *Actionable Knowledge Graph* (AKG) and briefly describe the related research task organized at NTCIR-13 (NII Testbeds and Community for Information access Research)[1] framework. AKG is considered as a specialized version of KG that contains data on the range of possible actions and affordances in relation to particular entity types and their instances. Automatically *constructing AKGs based on open information extraction* is then one important research objective. The other one relates to the problem of *optimizing the result pages for facilitating users' actions* and mainly consists of selecting most appropriate actionable interfaces for user queries that contain underlying actionable intent (e.g., buying, booking, downloading, comparing, creating). Our motivation is to allow researchers evaluate different approaches for AKG construction including statistical approaches, open information extraction, ontology-based methods to learn rules from kBs and others. With the standardized settings of the proposed task we can compare different approaches under the same conditions.

In this paper we make the following contributions: (1) *We provide a general overview of the research problem of automatically extracting actions relevant to input entities.* (2) *We discuss novel dedicated datasets for evaluating the entity-centric action retrieval constructed in the context of NTCIR-13 AKG task.*

2 Background

Actions are a fundamental component of AKG. For a given entity (e.g., an entity included in a user query) AKG should contain its relevant actions together with their related *descriptive complementary data* including constraints, actor types, temporal aspects and others.

In its basic form, an action is defined as an event composed of two parts: an *action form* and a *modifier*. The action form corresponds to the event as described by a verb or related PoS tags. The modifier is either an object of an action form or content that provides detailed context for the action form which provides important information on the character of action, purpose, situation, etc. For example, for the entity "`tokyo`" the examples of the relevant actions would be "`see modern architecture`" and "`learn japanese`" with "`modern architecture`" and "`japanese`" being modifiers. The entity "`SIGIR2017`" could have actions "`attend`" and "`learn IR technologies at tutorials`". Other examples can be found at AKG task website[2]. Note that an action is not constrained to the one that can be performed by a user (searcher). Further refinements can however filter out those actions that realistically cannot be completed by a searcher, either by utilizing searcher's profile and context (e.g., browsing history, location and demographics) or simply by assuming an average persona.

The above-mentioned descriptive data for an action embraces a range of components that enable more precise execution or realization of an action including constraints, actors, typical forms of action completion etc. Many of such components can be found in generic resources like VerbNet[3] or schema.org[4]. Of special importance are entity predicates that determine the character of an action that can be performed in relation to the entity. For example, for an action "`cook on the bbq or grill`" performed in relation to the entity "`goat meat`", entity's attributes like "`production date`", "`weight`" or "`brand`" are all relevant for performing the action[5].

3 AKG Task

In this section we describe the datasets developed for Actionable Knowledge Graph Task (AKG)[6] under NTCIR-13 framework. NTCIR (NII Testbeds and Community for Information access Research) is a series of workshops similar to TREC for evaluating technologies of information retrieval and access. AKG is composed of two subtasks: *Action Mining Subtask* (AM) and *Actionable Knowledge Graph Generation Subtask* (AKGG). AM requires returning relevant actions for input entities, while for AKGG participants need to submit relevant properties for the combination of entity and one of its actions.

Note that system descriptions, evaluation results and their detailed analysis as well as the details of settings used for gathering crowdsourcing annotations are to be provided in the task overview paper [3].

[1] `http://research.nii.ac.jp/ntcir/index-en.html`
[2] `http://ntcirakg.github.io/tasks.html`
[3] `https://verbs.colorado.edu/verb-index/`
[4] `http://schema.org`
[5] Other examples can be found at `http://ntcirakg.github.io/tasks.html` and in Tab. 3.
[6] `http://ntcirakg.github.io/`

3.1 Action Mining Subtask

The formal run dataset of AM task consists of 200 test entities sampled from a set of query log and question answering datasets. In particular, we grouped together the question answer and query data from Yahoo Webscope[7] and run an entity linker[4] over each question/query and selected the top-1 ranked entity. We then have selected entities based on their importance in the datasets estimated by the frequency of occurrence. Table 1 shows several examples of inputs that participants receive. For each such input, that is, in particular, for a given entity type (e.g., Product) and instance entity (e.g., "`Final Fantasy VIII`"), up to 100 potential actions that can be taken in relation to the entity (e.g., "`play on android`", "`buy new weapons`", "`learn junction system`" should be returned by participants. The actions are to be found by participants based on any data source they wish to use and any methodology. Several example relevant actions for the test instance marked by #1 in Tab. 1 are shown in Tab. 2). The format of each action form contains verb (e.g., "`play`") and modifier[8] (e.g., "`on Android`"). As semantics of actions can differ quite much depending on their modifiers, participants are allowed to submit up to three actions that share the same verb.

#	Entity	Entity Type(s)	Wikipedia URL
1	Final Fantasy VIII	Product	https://en.wikipedia.org/wiki/Final_Fantasy_VIII
2	Yo-Yo Ma	Person	https://en.wikipedia.org/wiki/Yo-Yo_Ma
3	Zambia	Place	https://en.wikipedia.org/wiki/Zambia
4	York University	Organization	https://en.wikipedia.org/wiki/York_University

Table 1: Example test instances of AM subtask.

The evaluation of the runs submitted by the participating teams was done in two assessment stages. First, verbs from the submitted actions were judged as for their relevance irrespectively of their modifiers. This was done using the CrowdFlower[9] crowdsourcing platform based on results pooled from all the participating teams. The total depth of the pool was 20. The second level of assessment involved the full actions (verbs+modifiers) such that only the actions judged as the most relevant in the first assessment (L3 score, described below) were considered. Again, the selected results were pooled with the cut-off value equal to 20.

For completing both the assessments, CrowdFlower workers had to choose from the following options:

L3 Some people, organizations or other subjects definitely have taken or will take this action for the entity

L2 This action has been or will be definitely taken by the entity

L1 This action can be relevant for the entity

L0 There is no relevance of the action to the entity

For the performance testing the average values of nDCG@10, nDCG@20, nERR@10 and nERR@20 were used for both levels of assessment.

Verb	Object
play	on android
buy	new weapons
learn	junction system
watch	videos of other players
compare	with other games

Table 2: Example results for the input given in test instance #1 of Table 1.

Ranked Properties
Agent
ServiceType
Result
Location
StartTime

Table 3: Example results for the input given in test instance #1 of Table 4.

3.2 Actionable Knowledge Graph Generation Subtask

The second subtask is related to detecting descriptive data: entity predicates that are relevant for performing the action. Knowing such predicates should be useful for search engines to offer direct interfaces for action

[7] `https://webscope.sandbox.yahoo.com/`
[8] The modifier's length is limited to 50 characters. Modifier can be also missing (NULL).
[9] `http://www.crowdflower.com`

completion. Table 4 shows example test instances consisting of a search query, entity included in that query, the types of the entity, and action. Participants were asked to rank entity properties (as demonstrated in the example shown in Table 3 which corresponds to the test instance #1 in Table. 4) based on their relevance to the query. To give a concrete case of how the returned properties could be utilized in real world scenarios, let us suppose that a user issues a query "`request funding`". One could then imagine a search engine with automatically generated links to facilitate the execution of the task (i.e. "`applying for funding`") by the user. Such links could be categorized into groups based on ranked properties of the action as indicated in Table. 3 offering useful pieces of information (e.g., ranked lists of relevant "`Agents`" which offer fundings) to initiate and carry on the action.

#	Query	Entity	Entity Type(s)	Action
1	request funding	funding	thing, action	request funding
2	kyoto budget travel	kyoto	thing, place	visit a temple
3	consequences of flood	flood	thing, event	live in a flood area
4	how to use google maps	google maps	thing, intangible, service	create a google maps mashup

Table 4: Example test instances of AKGG subtask.

The query (input) can be ambiguous as in realistic search queries, and participants need to return the ranked list of relevant entity properties. Properties to be ranked and returned were those defined as attributes of the entity type in schema.org[10] vocabulary. Participants could submit up to three runs with 20 being the maximum number of ranked attributes.

Actions in the test queries were first taken from the outcomes of the Action Mining (AM) Subtask which were judged as relevant by CrowdFlower workers. Then they were manually selected by the task organizers. For the total of 200 queries, half of them had modifiers and half were missing any modifiers.

Note that we effectively assume that the search queries contain underlying actionable intents (e.g., query "`consequences of flood`" is assumed to contain actionable intent of, for example, "`living in the areas impacted by a flood`"). In reality, searchers may of course have non-actionable intents behind their query strings. Disambiguating actionable intent vs. no actionable intent is left for future investigation.

4 Discussion

There are a number of open research questions left in relation to testing actionable graph generation methods. In this section we briefly discuss some of them.

4.1 Action Format

First, the format of actions can be made more specific. For example, modifiers can be further divided into smaller components which could have or can lack data for a particular instance action. More detailed action structure could allow finer testing of effective solutions and building more customized and adaptable interfaces.

Furthermore, actions could be further represented as RDF triples instead of plain strings. Another option would be to synchronize the actions with ones described in dedicated knowledge bases such as VerbNet[11].

4.2 Evaluation of Actions

The other one relates to the evaluation of actions in AM task which consist of verbs and modifiers. We assume that the possible actions with respect to an entity are exclusive or independent from each other and the relationships among the actions are not explicitly taken into account. The actions, however, could be similar to each other, could have a type-subtype or causal relationships and so on. In particular, as we have found, for some entities, the returned candidate actions form a hierarchy, as some actions are correlated and some are sub-concepts of others. It might be then more effective to consider deploying more refined evaluation measures (e.g., [9, 8]) where hierarchical information is considered.

[10] `http://schema.org`
[11] `https://verbs.colorado.edu/verb-index/`

4.3 Usage of Crowdsourcing

The third open research question is about the usage of crowdsourcing platforms. For generating high-quality candidate actions, a number of fundamental issues have to be addressed, such as named entity recognition and entity resolution. Most of the participants appeal to the off-the-shelf pipelines. For particular entities, it is possible that all the submitted runs fail to provide high-quality candidates. Low quality results may be obtained due to errors during the phases of, for example, natural language processing (NLP) and information extraction. In result, the final standard answers will be impacted. Moreover, another challenging issue is to alleviate the impact of inaccurate annotations by malicious workers. State-of-the-art practices for ensuring good quality of annotations should be implemented [1].

4.4 Interface Design

Another open research question relates to the interface design, which should allow users to access the information in an effective way. Towards this direction, exploratory search interfaces could be proposed based on the mined actionable information. Proposing and testing effective user interfaces is then another direction for the next evaluation tasks for AKGs.

5 Conclusions

Task oriented information retrieval is an emerging paradigm in search technologies. In this paper we have discussed the concept of Actionable Knowledge Graph and described the format of actions to be included in AKGs. We have then introduced the two subtasks proposed at the related NTCIR-13 AKG task which is designed for testing technologies aiming at extracting actionable components related to entities included in user queries as well as we have outlined the related test collections. The datasets created in relation to the NTCIR-13 AKG task can be obtained for research purposes.[12]

Future work can include categorization of queries based on the scope of their actionability, that is, the extent to which a searcher wishes to perform some action as well as deeper investigation of context elements that can support or lead to the successful execution of the actions. We plan also to investigate other aspects of the emerging paradigm of *task-oriented IR* which are related to Actionable Knowledge Graphs.

6 Acknowledgements

This research and development work was partially supported by the MIC/SCOPE #171507010.

References

[1] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar. Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, IEEE Computer Society 17(2):76–81, 2013.

[2] R. Blanco, B. B. Cambazoglu, P. Mika, and N. Torzec. *Entity Recommendations in Web Search*, pages 33–48. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[3] R. Blanco, H. Joho, A. Jatowt, H. Yu, and S. Yamamoto. Overview of ntcir-13 actionable knowledge graph (akg) task. In *Proceedings of NTCIR-13 Conference*, Tokyo, Japan, 2017.

[4] R. Blanco, G. Ottaviano, and E. Meij. Fast and space-efficient entity linking for queries. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15, pages 179–188, New York, NY, USA, 2015. ACM.

[5] A. Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.

[6] N. Dalvi, R. Kumar, B. Pang, R. Ramakrishnan, A. Tomkins, P. Bohannon, S. Keerthi, and S. Merugu. A web of concepts. In *Proceedings of the Twenty-eighth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '09, pages 1–12, New York, NY, USA, 2009. ACM.

[12] http://research.nii.ac.jp/ntcir/index-en.html

[7] T. Lin, P. Pantel, M. Gamon, A. Kannan, and A. Fuxman. Active objects: Actions for entity-centric search. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 589–598, New York, NY, USA, 2012. ACM.

[8] X. Wang, Z. Dou, T. Sakai, and J. Wen. Evaluating search result diversity using intent hierarchies. In *Proceedings of the 39th SIGIR*, pages 415–424, 2016.

[9] H. Yu, A. Jatowt, R. Blanco, H. Joho, and J. Jose. An in-depth study on diversity evaluation: the importance of intrinsic diversity. *Information Processing and Management*, 53(4):799–813, 2017.