Report on the Workshop on Personal Knowledge Graphs (PKG 2021) at AKBC 2021

Krisztian Balog
University of Stavanger
Norway
krisztian.balog@uis.no

Paramita Mirza
Max Planck Institute for Informatics
Germany
paramita@mpi-inf.mpg.de

Martin G. Skjæveland
University of Oslo
Norway
martige@ifi.uio.no

Zhilin Wang
University of Washington
USA
zhilinw@uw.edu

Abstract

The term *personal knowledge graph* (PKG) has been broadly used to refer to structured representation of information about a given user, primarily in the form of entities that are personally related to the user. The potential of personal knowledge graphs as a means of managing and organizing personal data, as well as a source of background knowledge for personalizing downstream services, has recently gained increasing attention from researchers in multiple fields, including that of Information Retrieval, Natural Language Processing, and the Semantic Web. The goal of the PKG'21 workshop was to create a forum for researchers and practitioners from diverse areas to present and discuss methods, tools, techniques, and experiences related to the construction and use of personal knowledge graphs, identify open questions, and create a shared research agenda. It successfully brought about a diverse workshop program, comprising an invited keynote, paper presentations, and breakout discussions, as a half-day event at the 3rd Automated Knowledge Base Construction (AKBC'21) conference. The workshop demonstrated that while the concept and research field of personal knowledge graphs is still in its early stages, there are many promising avenues of future development and research that already, and independently, have attracted the interest of several different communities.

Date: October 8, 2021.

Website: https://pkgs.ws.

1 Introduction

The concept of *personal knowledge graphs* (PKGs) has been around for a while, in recognition of the need to represent structured information about entities that are personally related to a user [Balog and Kenter, 2019]. Recent years have seen a growing body of research around PKGs, from extracting personal information for building/populating PKGs [Tigunova et al., 2019, 2020;

Yu et al., 2020; Montoya et al., 2018; Safavi et al., 2020; Vannur et al., 2020] to various ways of utilizing this information [Torbati et al., 2021; Gerritse et al., 2020; Kalokyri et al., 2018]. However, several open questions remain around the definition, construction, utilization, and practical realization of PKGs. Another key observation is that research is scattered across different communities, including that of Information Retrieval, Natural Language Processing, and the Semantic Web. The main aim of the workshop was to bring together researchers and practitioners from diverse areas to identify and discuss these open questions and make progress towards a shared research agenda.

The PKG'21 workshop was a half-day event, held at the 3rd Automated Knowledge Base Construction (AKBC'21) conference in a virtual format (due to the ongoing COVID-19 pandemic). AKBC aims to bring together researchers from diverse areas, which is why we chose that as our venue, as opposed to any individual community's conference.

The workshop solicited submissions on the following themes:

- Modeling personal knowledge (ontologies and knowledge representation);
- Populating and maintaining personal knowledge graphs (natural language processing, entity and relation extraction, information integration);
- Applications of personal knowledge graphs (including but not limited to intelligent personal assistants, personal information management systems, health information systems, recommender systems);
- Evaluation (evaluation measures and methodology, benchmark construction);
- Systems and toolkits.

We invited submissions of regular papers, position papers, demonstrators as well as encore talks (featuring work that has already been published in a leading conference or journal). This latter submission category turned out to be indeed the most popular one.

Overall, the workshop programme featured a keynote talk (by Cathal Gurrin), one regular paper, nine encore papers, and breakout group discussions.

2 Keynote

The keynote talk of the workshop was given by Cathal Gurrin (Dublin City University) on "Lifelogs as Personal Data Information Systems."

Lifelogging refers to phenomenon of people digitally recording their own daily lives, using various software and hardware sensors [Gurrin et al., 2014]. Driven by low-cost sensors, years of data can now fit on a $100 hard disk. The challenge for information scientists then is to facilitate effective use of this data in daily life, by developing useful tools for personal information access. Use-cases include augmenting health, wellbeing, memory, and cognition. Among these, Cathal identified the potential for memory support, where the lifelog works in synergy with our own memory, as one of the most interesting aspects.

One main research thread concern the organization, indexing, and enrichment of personal data. Machine learning methods generally work well for extracting information from multimodal data, such as finding objects and activities. There also lies a lot of opportunities in combining different sensing modalities to develop human activity models [Kahneman et al., 2004]. Personal knowledge graphs could potentially play a key role here, for organizing and enriching personal data. One main difference from general KGs is the curation of the data, which only the individual can do in this case. The incentives for doing that, and the added value this would bring, are nevertheless open questions.

Another thread consists of querying personal data, and in particular moving from event-based browsing interfaces to search engines [Doherty et al., 2012]. User queries are typically of the known-item search type (*"Find the moment when I was taking a photo of an A380"*), but there is additionally a strong emphasis on real-time and interactive search. Cathal spearheaded collaborative benchmarking activities in this domain, which include the NTCIR Lifelog track (2012–2014), to address the retrieval challenges, ImageCLEF (2017–2018), to build better annotation tools for visual and multimodal content, and the Lifelog Search Challenge (LSC, 2018–2021), for interactive retrieval. Specifically, the NTCIR 2014 Lifelog benchmark [Gurrin et al., 2019] consists of rich lifelog data for 60 days for 3 people (fully anonymized) and features three sub-tasks: lifelog semantic access (a search engine for the individual), lifelog insights (health related insights based on biometrics and images), and lifelog activity detection (finding the activities from an ontology of life activities). The ImageCLEF 2018 Lifelog task [Ionescu et al., 2018] has two sub-tasks: lifelog moment retrieval (retrieving specific moments, e.g., *"when I was shopping for wine in the supermarket"*) and activities of daily living understanding (summarizing lifelog data according to specific requirements, e.g., *"preparing meals (include making tea or coffee)"*). The dataset consists of 50 days of multimodal data from a lifelogger. Finally, the annual Lifelog Search Challenge[1] involves an interactive search engine challenge based on the NTCIR Lifelog dataset. Teams are scored based on the time taken to find answers to known-item search topics, using their own prototype interactive lifelog search engines.

Clearly, sharing data is difficult, but the above benchmarks attest to progress in this regard, even if datasets comprise only a few individuals. The ethics of how to use rich personal data and doing so in a privacy-aware manner remains an open issue.

3 Papers and Presentations

The PKG'21 workshop committee accepted one regular paper and nine encore talks for presentation. Summaries of each can be found below (presenters' names are typeset in boldface).

Personal Health Knowledge Graph for Clinically Relevant Diet Recommendations (regular paper [Seneviratne et al., 2021])
Oshani Seneviratne, *Jon Harris, Ching-Hua Chen and Deborah McGuinness*

Oshani demonstrated how patients with Type 2 Diabetes (T2D) can get personalized diet recommendations consistent with clinical guidelines by exploiting semantic modeling and reasoning on the patient's own health records and food log together with an OWL ontology of clinical guidelines. The patient's health records and food logs are represented in a Personal Health Knowledge Graph (PHKG), modeled according to the Personal Health Ontology for capturing lifestyle behaviors related to food consumption, and using the FoodKG ontology for representing food recommenda-

[1] http://lsc.dcu.ie/

tions. The Personal Health Ontology was developed for this task through interviews with patients diagnosed with T2D and is used for representing personal data such as food likes and dislikes, and food, cooking and restaurant recommendations. The FoodKG, created by the authors in previous work, is used for representing more general, non-personal facts about food such as food recipes. The dietary guidelines ontology was modeled for this task taking guidelines from the American Diabetes Association and representing each guideline with the necessary and sufficient conditions for guideline compliance, plus the action to take in case on non-compliance. A Time Series Summarization (TSS) framework [Harris et al., 2021] was extended to generate RDF triples (rather than natural language summaries) from food logs for the PHKGs. The generated PHKGs were evaluated using a competency question-based methodology, testing if the PHKG can answer questions like *"Have I been following a Mediterranean diet?"* and *"What should I eat for breakfast?"*. The resources developed in this work are available.[2]

Data Augmentation for Personal Knowledge Base Population (encore talk [Vannur et al., 2020])
*Lingraj S Vannur, **Balaji Ganesan**, Lokesh Nagalapatti, Hima Patel, and MN Thippeswamy*

Balaji showed how the various techniques (named entity recognition, relation extraction, entity resolution, entity type classification, slot filling, link prediction between people) involved in knowledge base population can be adapted for personal knowledge graph population. Specifically, it uses certain relations relevant to personal knowledge graphs from the TACRED dataset [Zhang et al., 2017] to perform such experiments. Building upon this set of techniques, Balaji showed how rule-based data augmentation can improve their performance. These data augmentation techniques make use of dictionaries to identify synonyms of verbs for relation extraction and entity types for augmentation of entities. With these methods, the recall for knowledge base population is substantially improved, while precision is maintained at a high level. Together, the proposed techniques represent a promising approach to personal knowledge base population, especially in real world settings where the cold start problem presents challenges.

Extracting Personal Information from Conversations (encore talk [Tigunova et al., 2019, 2020])
***Anna Tigunova**, Andrew Yates, Paramita Mirza, and Gerhard Weikum*

Two related papers were presented in this talk by Anna Tigunova. She explained how they collected a dataset from Reddit that contains information on personal attributes such as users' professions and hobbies. These personal attributes can be challenging because they have a wide range of possibilities, as opposed to other personal attributes with a closed set of options (e.g., age group or gender). They collected this dataset using a weakly supervised manner, based on whether users expressed explicit personal assertions such as *"I work as a doctor."* Using this dataset, they predict user attributes using their utterances on Reddit. Not only did they use traditional IR methods (e.g., BM25 and TextRank) and machine-learning-based methods (e.g., BERT), they also sought to improve the zero-shot transfer-ability of models. This is done by training a model to identify keywords that would suggest such attributes, meaning that previously unseen keywords

[2] https://semantics-for-personal-health.github.io

associated with a personal attribute can also be used to support predictions. Another advantage of this method is that predictions are highly interpretable. Together, they present a promising direction in collecting a weakly supervised dataset for modeling personal attribute prediction.

Bias in Conversational Search: The Double-Edged Sword of the Personalized Knowledge Graph (encore talk [Gerritse et al., 2020])
Emma J. Gerritse*, Faegheh Hasibi, and Arjen P. de Vries*

Emma discussed the prospects of utilizing personal knowledge graphs to tailor responses in conversational AI systems. In particular, it raises potential sources of biases arising from such personalization including bias from knowledge graph construction, search algorithms as well as user demographic stereotypes. Furthermore, biases can arise from the intersection between people and algorithms, requiring consideration on the degree of personalization by algorithms to the information supplied. Emma also discussed a few approaches to manage such potential biases. Conversational systems can ignore potential biases, which Emma suggests as impractical given the extent of personalization in conversational search systems. More suitable approaches are putting users in control of the conversation flow in case of detected biases and allowing both the user and system to jointly reflect upon such biases. Overall, this paper raises important issues surrounding detecting and mitigating potential biases in personalizing conversational search systems.

PROV4ITDaTa: Transparent and direct transfer of personal data to personal stores (encore talk [De Mulder et al., 2021])
Gertjan De Mulder*, Ben De Meester, Pieter Heyvaert, Ruben Taelman, Ruben Verborgh, and Anastasia Dimou*

Gertjan demonstrated PROV4ITDaTa, a web application that allows users to transfer personal data into an interoperable format stored in their personal data store, using the open-source solutions RML.io,[3] Comunica,[4] and Solid.[5] The RML.io toolset is used to describe how to access data from different service providers and generate interoperable datasets, Comunica is used to query these dataset and to generate enriched datasets, while Solid Pods store the generated data as RDF Linked Data in personal data stores. PROV4ITDaTa is transparent in that each component of the pipeline can be configured and may automatically generate detailed provenance trails. One of the main motivations behind PROV4ITDaTa is to facilitate a shift towards interoperable solutions based on open-source tools and open standards where users have control of their data, and providers can focus on their services instead of trying to adhere to interoperability requirements.

[3] https://rml.io/
[4] https://comunica.dev/
[5] https://solidproject.org/

A Knowledge Base for Personal Information Management (encore talk [Montoya et al., 2018])
David Montoya, *Thomas Pellissier Tanon, Serge Abiteboul, Pierre Senellart, and Fabian M. Suchanek*

David presented Thymeflow, an open-source framework for integrating personal user data, such as email messages, calendar data, contacts, and location history, from different sources to a single coherent RDF knowledge base. The knowledge base data is represented using an extension of the schema.org vocabulary, and is made available for querying through a SPARQL interface. Thymeflow is a personal web application that can pull data from various sources as set up by the user. The data is converted from its original formats into the personal knowledge base, hence storing a converted copy of the original data. Central to the architecture of the system are its synchronisers and enrichers. Synchronisers are responsible for updating or deleting local data if the original sources are changed. Enrichers perform inference tasks on loaded data such as entity resolution, event geolocation, and other knowledge enrichment tasks. Having a single source of personal information allows the user to query within and across different dimensions on their own data, and to perform analytics over emails, events, and locations.

Dialogue-Based Relation Extraction (encore talk [Yu et al., 2020])
*Dian Yu, **Kai Sun**, Claire Cardie, and Dong Yu*

Kai presented the DialogRE dataset,[6] which is based on multi-turn dialogues from the TV series Friends. Based on such dialogues, the task is to identify relations involving characters in the dialogue. For instance, it seeks to identify whether two characters are friends, the school that a character attended, and their impression of another character. Based on this task, the authors found that a character-aware model, which appends the name of the speaker before their utterance can improve modeling performance. This suggests that including speaker names can implicitly capture knowledge about a speaker, thereby improving its performance in the corresponding relation extraction task. Together, the dataset and model presented in this paper supports further research in extracting relations involving characters from a dialogue corpus.

Toward Activity Discovery in the Personal Web (encore talk [Safavi et al., 2020])
Tara Safavi, *Adam Fourney, Robert Sim, Marcin Juraszek, Shane Williams, Ned Friend, Danai Koutra, and Paul N. Bennett*

Tara presented their work on how individuals' personal information collections in the form of emails, files, appointments, web searches, and contacts can be effectively used to support applications that depend on an understanding of activities. To do so, they recorded two days of activities on participants' main work computers and used such activities to make recommendations of activities. Using this dataset, they represent various entities involved in these activities as nodes and common mention of entities within an activity as relations. A graph learning algorithm was then used to learn embedding representations of such nodes, which were initialized using the noun-phrases found in the node, weighted by term-frequencies or a Latent Semantic

[6] `https://dataset.org/dialogre/`

Analysis representation of such noun-phrases. Such an algorithm was found to be not only orders of magnitude faster than baselines but also supported downstream applications more effectively, both in terms of intrinsic metrics (e.g., how related two nodes are as rated by participants) and extrinsic metrics (e.g., usefulness of recommendations made using these node embeddings).

You Get What You Chat: Using Conversations to Personalize Search-Based Recommendations (encore talk [Torbati et al., 2021])
Ghazaleh H. Torbati*, Andrew Yates, and Gerhard Weikum*

Ghazaleh discussed how information extracted from conversations may be used to personalize search-based recommendations concerning books, food recipes, and travel destinations. Prior works typically use information that has explicitly been supplied by the user, for instance clicks, likes, and ratings on questionnaires. However, natural conversations between people can be a rich source of information that has been untapped, with distinct advantages over user-supplied information such as not having to update interests as they evolve over time. They approach this task by collecting a dataset of two people speaking to each other about specific interests concerning books, food, and travelling. Information is extracted from such conversations using named entity recognition and disambiguation, then the identified entities are re-ranked to provide personalized recommendations using a variety of methods, including BM25, language modeling, and neural re-ranking. The authors compare this approach with personalization using questionnaire-collected data (on demographics, personality, and general interests in books, food and travel). They find that the two approaches are on par with each other, but have their respective advantages in certain areas. This means that personalization of search-based recommendations can benefit from information expressed during conversations.

YourDigitalSelf: A Personal Digital Trace Integration Tool (encore talk [Kalokyri et al., 2018])
Varvara Kalokyri*, Alexander Borgida, and Amélie Marian*

Varvara presented the YourDigitalSelf system that allows individuals to manage, make inferences, and personalize recommendations using heterogeneous digital traces. This tackles the challenge that personal information is usually fragmented across various sources including Facebook locations, Amazon purchase histories, Yelp reviews, email messages, social media messages, and documents on online disk drives. To consolidate the various sources of information, they created a unified schema to organize personal information from various sources. This schema is based on W5H (What, When, Who, Where, Why, How), which captures the universal attributes that activities have. In addition, the system also accounts for the strength of evidence for particular events. For instance, a discussion for dinner is weak evidence (that a dinner has happened) while a restaurant reservation or a Facebook post tagging people at a restaurant are strong evidence. Furthermore, the system used named entity recognition techniques to classify named entities found in data sources, to produce meaningful semantic groupings (such as restaurants) for users to view. Finally, to maintain privacy of personal information, all personal data is kept on the user's personal Android device. Together, this system represents a first step to supporting users to effectively utilize diverse sources of their online digital traces.

4 Breakout Discussion

The workshop included a breakout session, where participants split into two groups, based on their preferences and interests. The first group discussed topics concerning the *creation* of personal knowledge graphs, while the second group discussed the *use* of personal knowledge graphs.

4.1 Creating Personal Knowledge Graphs

The breakout group consisted of Anna Tigunova, Zhilin Wang, Paramita Mirza, and Martin G. Skjæveland. The group identified and discussed different topics related to the creation of personal knowledge graphs, primarily focusing on relevant research challenges. The outcome of the discussion can be categorized as follows:

Characterization What characterizes a PKG, and how does it differ from a regular KG? What does "personal" in PKG mean? It could be taken to mean (objective) facts about the user (*I ate lunch at restaurant X on date Y. I like fish.*), subjective beliefs of the user (*[I believe that] Pineapple pizza is just wrong. The Earth is flat.*), or objective facts that are of particular interest to the user (*Pineapple pizza is also often called Hawaiian Pizza*). The context in which the data was created, such as the time and place, is important for understanding and later correctly using the data recorded in a PKG.

Representation How should one formally and technically represent the PKG, and what kind of computational properties should the PKG be able to support? What kind of knowledge representation framework or ontology language is best suited for the task? The expressivity of the formal representation language for the PKG should match its wanted characteristics. The expressivity should also be balanced with the complexity of important computational tasks. Finally, the technology used should preferably be widely available and suitable for the task at hand.

Integration What kinds of data sources (from unstructured, to natural language, semi-structured, and structured) are suitable for constructing a PKG and what types of methods and tools are required for data extraction, comprehension and representation? What kind of external sources are relevant for integrating or interfacing with the PKG data? How is external generic knowledge graphs facts that may be in conflict with the PKG data treated? For example, can beliefs in the PKG (*The Earth is flat*) co-exist with conflicting facts (*The shape of the Earth is a oblate spheroid*) in external KGs like DBPedia.org or Wikidata.org?

Maintenance The data in the PKG must support updates, both in terms of error corrections and in terms of data that naturally change over time. How does one support temporal changing or historical data?

4.2 Using Personal Knowledge Graphs

The second discussion group, consisting of Cathal Gurrin, David Montoya, Gabriel Maia, Ghazaleh H. Torbati, Krisztian Balog, Oshani Seneviratne, and Vimig Socrates, discussed how personal knowledge graphs may be utilized. As applications are inherently data driven, access to personal data would open up avenues for a whole new dimension of functionality. This, however, would need to be done such that the privacy of users is ensured and they remain in control and ownership of their own data. There is a broad range of application scenarios that would benefit from access to PKG, including but not limited to search and recommender systems, education and learning, electronic health records, and digital personal assistants. More broadly, access to structured personal data would give rise to new tasks, architectures, and workflows/protocols.

The biggest challenge and obstacle, not surprisingly, is to get people to share their data. There are many open questions (and risks) here around privacy and legally compliant ways of using (and releasing) data. Still, even if proper safeguarding mechanisms were in place, we seem to be missing the "killer app"—something so valuable or necessary that it would incentivise people to share their personal data.

As different applications have different "data needs," one challenge is the identification and curation of useful data sources for each. At the same time, it is a desired property of a PKG to be as broad as possible, capturing a holistic picture of the user. Finally, personal knowledge graphs represent "small data," and as such likely calls for tailored solutions.

5 Summary

The PKG'21 workshop was designed to create a forum for, and successfully brought together, researchers and practitioners with a diverse background, who are interested in the concept, development, and utilization of personal knowledge graphs. These are reflected in a diverse workshop program comprising an invited keynote and paper presentations from various fields, including Information Retrieval, Natural Language Processing, and the Semantic Web.

The great potential of personal knowledge graphs as a means of managing and organizing personal data, which can then be utilized in downstream services, is largely agreed. However, the very definition of a PKG has not crystallized yet: Does the term "personal" refer to objective facts about the user, subjective beliefs of the user, or to objective facts that are of particular interest to the user? The answer to this question can most likely only be given with respect to a given application context. Furthermore, the biggest challenge in pursuing research around PKGs is ensuring users' privacy in gathering, utilizing, and sharing personal data. Establishing a research community for this particular topic is one way to overcome such a challenge, thus, we especially look forward to data creation and sharing efforts that help push the agenda on PKGs further.

Acknowledgments

We thank all speakers and workshop participants for their contributions. We would also like to thank our program committee members: Faegheh Hasibi (Radboud University, Netherlands), Kjetil Kjernsmo (Inrupt, Norway), Filip Radlinski (Google, London), Anna Tigunova (Max Planck Institute for Informatics, Germany), Ruben Verborgh (Ghent University, Belgium), Gerhard Weikum (Max Planck Institute for Informatics, Germany), and Charles Welch (University of Michigan, USA).

References

Krisztian Balog and Tom Kenter. Personal knowledge graphs: A research agenda. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR '19, pages 217–220, 2019.

Gertjan De Mulder, Ben De Meester, Pieter Heyvaert, Ruben Taelman, Anastasia Dimou, and Ruben Verborgh. PROV4ITDaTa: Transparent and direct transferof personal data to personal stores. In *Companion Proceedings of the Web Conference 2021*, WWW '21, pages 695–697, 2021.

Aiden R. Doherty, Katalin Pauly-Takacs, Niamh Caprani, Cathal Gurrin, Chris J. A. Moulin, Noel E. O'Connor, and Alan F. Smeaton. Experiences of aiding autobiographical memory using the sensecam. *Human–Computer Interaction*, 27(1–2):151–174, 2012.

Emma J. Gerritse, Faegheh Hasibi, and Arjen P. de Vries. Bias in conversational search: The double-edged sword of the personalized knowledge graph. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, ICTIR '20, pages 133–136, 2020.

Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. Lifelogging: Personal big data. *Found. Trends Inf. Retr.*, 8(1):1–125, jun 2014.

Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Van-Tu Ninh, Tu-Khiem Le, Rami Albatal, Duc-Tien Dang-Nguyen, and Graham Healy. Advances in lifelog data organisation and retrieval at the NTCIR-14 Lifelog-3 task. In *NII Conference on Testbeds and Community for Information Access Research*, NTCIR '19, pages 16–28, 2019.

Jonathan J. Harris, Ching-Hua Chen, and Mohammed J. Zaki. A framework for generating summaries from temporal personal health data. *ACM Trans. Comput. Healthcare*, 2(3), jul 2021.

Bogdan Ionescu, Henning Müller, Mauricio Villegas, Alba García Seco de Herrera, Carsten Eickhoff, Vincent Andrearczyk, Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Sadid A. Hasan, Yuan Ling, Oladimeji Farri, Joey Liu, Matthew Lungren, Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, and Cathal Gurrin. Overview of ImageCLEF 2018: Challenges, datasets and evaluation. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, CLEF '18, pages 309–334, 2018.

Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702):1776–1780, 2004.

Varvara Kalokyri, Alexander Borgida, and Amélie Marian. Yourdigitalself: A personal digital trace integration tool. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 1963–1966, 2018.

David Montoya, Thomas Pellissier Tanon, Serge Abiteboul, Pierre Senellart, and Fabian M. Suchanek. A knowledge base for personal information management. In *Linked Open Data Workshop at the World Wide Web Conference*, LDOW '18, 2018.

Tara Safavi, Adam Fourney, Robert Sim, Marcin Juraszek, Shane Williams, Ned Friend, Danai Koutra, and Paul N. Bennett. Toward activity discovery in the personal web. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, WSDM '20, pages 492–500, 2020.

Oshani Seneviratne, Jonathan Harris, Ching-Hua Chen, and Deborah L. McGuinness. Personal health knowledge graph for clinically relevant diet recommendations. *arXiv*, cs.HC/2110.10131, 2021.

Anna Tigunova, Andrew Yates, Paramita Mirza, and Gerhard Weikum. Listening between the lines: Learning personal attributes from conversations. In *The World Wide Web Conference*, WWW '19, pages 1818–1828, 2019.

Anna Tigunova, Andrew Yates, Paramita Mirza, and Gerhard Weikum. CHARM: Inferring personal attributes from conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, EMNLP '20, pages 5391–5404, 2020.

Ghazaleh H. Torbati, Andrew Yates, and Gerhard Weikum. You get what you chat: Using conversations to personalize search-based recommendations. In *Proceedings of the 43rd European Conference on IR Research*, ECIR '21, pages 207–223, 2021.

Lingraj S Vannur, Balaji Ganesan, Lokesh Nagalapatti, Hima Patel, and MN Thippeswamy. Data augmentation for personal knowledge base population. *arXiv*, cs.IR/2002.10943, 2020.

Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. Dialogue-based relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL '20, pages 4927–4940, 2020.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, EMNLP '17, pages 35–45, 2017.