Using Semantic-based Adaptive Relevance Prediction to Enhance Entity Recommendation for Personal Knowledge Assistance

Mahta Bakhshizadeh[1,2,*], Heiko Maus[1] and Andreas Dengel[1,2]

[1]*German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany*

[2]*University of Kaiserslautern-Landau (RPTU), Kaiserslautern, Germany*

Abstract

Personal knowledge assistance tools are designed to support knowledge work by delivering contextually relevant information and recommendations, thereby enhancing productivity and decision-making. Entity recommendation is a form of knowledge assistance that suggests relevant entities commonly sourced from public knowledge bases, like DBpedia, based on user context to improve productivity in daily digital tasks. In this study, we explore which similarity metrics within RDF2Vec knowledge graph embedding are most effective at capturing users' personal interpretations of entity similarities within their specific contexts. Accordingly, we propose a semantic-based recommendation method that includes an adaptive relevance prediction module to dynamically evaluate entity relevance by incorporating user feedback. Our approach is benchmarked on RLKWiC, a publicly available dataset of real-life knowledge work in context, and demonstrated a twenty percent improvement over the established baseline for entity recommendation, highlighting its potential to enhance knowledge work support.

1. Introduction

Recommender systems (RS) have become a major technology in a wide range of applications, from e-commerce and social media to digital entertainment [1]. Traditional RS primarily rely on collaborative filtering methods [2], which leverage user-item interaction data, often enhanced by machine learning techniques, to predict user preferences. However, these methods typically do not utilize the vast amounts of structured and unstructured knowledge available about the domain of interest. This gap has led to the emergence of Knowledge-aware RS (KaRS), which aim to integrate domain-specific knowledge into the RS to improve not just the accuracy but also the relevance and interpretability of recommendations.

KaRS extend beyond the conventional data-driven approaches by incorporating rich semantic information from various knowledge sources, such as ontologies, Knowledge Graphs (KGs), and other structured databases. These systems leverage this knowledge to provide more contextually relevant recommendations, allowing them to address some of the limitations of traditional recommenders, such as cold start problems and lack of explainability. By using knowledge bases and KGs, KaRS can infer new relationships between items and users, capture deeper insights about user preferences, and understand the semantics behind user interactions [3].

The integration of knowledge sources into RS represents a shift towards a more comprehensive approach, where the goal is not only to predict what a user might like but also to provide recommendations that are contextually appropriate and semantically meaningful. As such, knowledge-aware and conversational RS are at the forefront of advancing the field by leveraging the semantic richness of knowledge bases and the interactive nature of conversational AI, ultimately enhancing user satisfaction and engagement [4].

Over the past two decades, a limited number of studies have investigated the integration of KaRS with Personal Knowledge Assistance (PKA). This emerging application seeks to develop RS that can continuously provide Knowledge Workers (KW) with the most relevant and useful information based on their specific context, thereby enhancing productivity in their daily tasks [5]. KW, including professionals such as architects, engineers, scientists, lawyers, and academics, rely heavily on knowledge as their primary asset [6] and are primarily focused on processing and applying information rather than engaging in manual labor [7].

In a relevant study, KG-based RS were identified as one of three promising types of RS capable of addressing the complexities of this research goal [5]. A knowledge-aware recommender system integrated into the work environment can support KW in various ways throughout their daily activities. For instance, during the process of writing a paper, such systems could recommend relevant articles, suggest contacts for consultation based on previous research and experience, propose appropriate research tools, and highlight upcoming conferences. These recommendations can be derived from different layers of the KW's information space (personal, corporate, and global) tailored to both their immediate needs and long-term preferences [8].

While some studies have concentrated on exploring the personal and corporate information spaces of KW for recommendation, often framing it as proactive information delivery [9] or information re-finding [10], other research has focused on leveraging public knowledge bases for recommendations. This latter approach is commonly referred to as Entity Recommendation (ER), where entities typically refer to resources within public KGs, such as DBpedia[1] [11, 12, 13].

The goal of our study is to explore how incorporating user feedback into ER systems can enhance their accuracy in suggesting relevant entities to KW, based on the users' context. Specifically, we aim to investigate whether a semantic approach, which adapts recommendations by measuring the distance between entities in embedding space, can improve the relevance of these suggestions. This raises an additional research question: Are the representations of entities from public KGs, generated through common embedding methods, aligned with users' personal interpretations of entity relevance within their self-defined contexts? Our study seeks to provide insights into these questions, assessing the potential of adaptive, feedback-driven systems in refining entity recommendations.

In the subsequent section, we discuss the evolution of ER towards PKA followed by an introduction to a publicly available benchmark designed for evaluating ER in PKA. We then present our proposed approach, which uses semantic similarities to dynamically predict the relevance of entities, thereby enhancing ER performance on the established benchmark. The paper is then concluded along with a brief overview of potential future research directions.

2. Personal Knowledge Assistance through Entity Recommendation

There has been a significant evolution of ER in the web search domain [14, 15] with extensive benchmarking using popular datasets [16] such as Movielens [17]. While PKA has not gained the same level of popularity as these domains, it has nonetheless seen notable contributions and advancements. This evolution reflects a shift from basic, application-specific models to sophisticated, context-aware systems that understand and anticipate user needs across various digital environments.

Initially, ER systems relied on limited user inputs and predefined logs, such as query histories or browsing data. For instance, the proactive information retrieval efforts focused on screen surveillance and the use of optical character recognition to analyze all content on a user's screen, enhancing task detection accuracy and proactive retrieval through digital footprints [18, 19]. An earlier study explored ER application within email management, as demonstrated by the SmartOffice extension for Microsoft Outlook [20]. This tool integrated email processing with enterprise workflows, enhancing the efficiency of handling process-relevant emails and documents. Evaluation results highlighted SmartOffice's potential to significantly improve workflow integration and user acceptance, showcasing the early promise of ER in supporting professional tasks through email.

---

[1]https://www.dbpedia.org

As research progressed, there was a greater focus on user intentions, task goals, and the factors driving information search behaviors. Studies highlighted that search behaviors were often linked to creative processes triggered by prior digital activities, underscoring the need for contextual factors in RS design [21]. This led to the development of entity-based systems that deliver actionable recommendations across multiple applications by continuously monitoring digital activities and capturing context through screen frames. Such systems, exemplified by EntityBot, offer recommendations without explicit queries, enhancing user productivity and satisfaction by reducing cognitive load [22, 12, 11].

Further advancements involved refining the methodologies for capturing and using contextual information. Techniques like Dirichlet–Hawkes processes were used to model context from digital traces, enhancing web search query augmentation based on comprehensive user activity data [23]. Additionally, integrating multimodal data from spoken conversations and digital activities improved task predictions and ER, thereby supporting users more effectively in digital environments [24].

Entity footprinting, which models contextual user states through digital monitoring and uses latent representations of entities to predict relevance, demonstrated improved accuracy in user state detection and entity prediction [25]. Most recently, research utilizing transformer architectures to model digital activity contexts has shown promise in predicting personalized information needs for various tasks, suggesting that a broader use of contextual data can enhance the effectiveness of RS [26].

The evolution of ER systems towards PKA has emphasized contextual awareness, multimodal data integration, and proactive support, reflecting a continuous effort to improve the relevance and utility of recommendations in meeting users' information needs. However, despite the aforementioned valuable contributions, a significant barrier to advancing the use of RS for supporting KW has been the lack of a standardized framework for evaluating and benchmarking these methods [27]. Most of the conducted experiments rely on proprietary datasets that are, at best, partially available, leading to challenges in reproducing results and making fair comparisons across different approaches [28]. The RLKWiC Benchmarking Dataset for ER[2], that is introduced in the following section, fills this gap by extending a publicly available dataset[3] towards promoting transparency and comparability, enabling researchers to more effectively evaluate and contrast various methods [13].

3. Entity Recommendation on RLKWiC

RLKWiC is a publicly available dataset of Real-Life Knowledge Work in Context, gathered by monitoring computer interactions of 8 volunteers over 2 months. It aims to provide a standardized benchmark for evaluating PKA services by offering multiple information dimensions, including detailed user contexts, documents, semantics, events, and sessions [28].

This dataset is extended to create a community benchmark by simulating an ER scenario where participants were given entities extracted from selected segments of their captured activities across various contexts [13]. In this setup, 1,850 entity recommendations were simulated across 56 different contexts within the dataset. After eliminating duplicates, these entities were grouped by context and presented to participants for explicit feedback.

To evaluate the relevance of recommendations in relation to their respective contexts, participants were instructed to rate the recommended entities on a 3-point scale:

- **Irrelevant (0):** Indicates that there is no relevance between the recommended entity and the corresponding context.
- **Relevant (1):** Suggests that there is a connection between the entity and the context, but it does not fully encapsulate or represent the context.
- **Representative (2):** Denotes that the entity is closely aligned with the context, reflecting a high degree of relevance, such that the context can be inferred to be specifically about this entity.
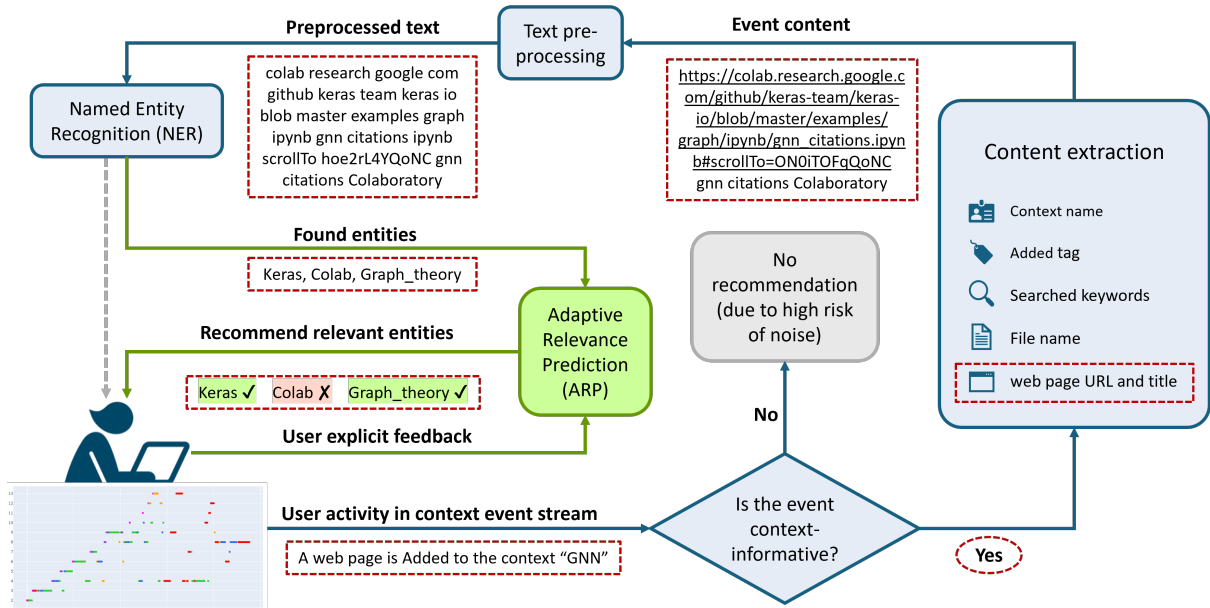

Figure 1: Incorporating Adaptive Relevance Prediction (indicated in green) into the entity recommendation scenario on RLKWiC.

It was noted that participants showed a preference for receiving information about entities labeled as representative in the corresponding context. For example, a participant who created the "GNN" context (a context about Graph Neural Networks), expressed interest in obtaining information about representative entities such as Message Passing[4] and Graph Representation[5] within this context. Conversely, participants generally found that explicit information about entities marked as relevant was less helpful in most scenarios. Nevertheless, these relevant entities can provide valuable, contextually rich information and enhance context representation learning for various information tasks. For example, in a job search context, the entity of Mannheim[6], a city in Germany where a participant was looking for a job, was labeled as relevant. Although providing direct information about Mannheim may not seem necessary in this context, utilizing this data indirectly—such as filtering job recommendations to only include positions in Mannheim—could lead to more pertinent results.

Participants could also suggest additional entities they considered relevant but were not included in the original recommendations. Combining participant feedback with these additional suggestions resulted in a dataset of over a thousand entities, each labeled with explicit relevance scores, enhancing the RLKWiC dataset.

This ER benchmark dataset provides comprehensive details for each recommendation case, including timestamps and participant-assigned scores. To establish a baseline for future research, the performance of a simulated ER scenario in recommending relevant and representative entities is also reported. With this scored entity set for each context, the challenge is to develop a recommendation strategy that can be simulated on the RLKWiC dataset to maximize the number of relevant and representative entities while minimizing irrelevant ones [13].

In the following section, we demonstrate how incorporating an Adaptive Relevance Prediction (ARP) module into the simulated ER scenario has allowed us to outperform the established baseline.

4. Adaptive Relevance Prediction Using Semantic Similarities

In the previously simulated ER scenario on RLKWiC (illustrated in Figure 1), participants were provided with entities derived from selected portions of their captured activities across various contexts.

---

[4]http://dbpedia.org/resource/Message_passing

[5]http://dbpedia.org/resource/Graph_representation

[6]http://dbpedia.org/page/Mannheim

In this scenario, each recorded event was first classified as either context-informative or not by the ER system. Context-informative events, which have evidence of the user's implicit confirmation of relevance include five types of activities: naming a newly defined context, making a search query, and adding a tag, file, or web page to the current context. Given the abundance of activities within each context, including considerable noise (e.g., distractions from irrelevant emails), only the context-informative events were considered as recommendation triggers. In the case of any of the mentioned triggering events, the corresponding content was extracted and pre-processed for the next step. This involves removing symbols and irrelevant strings (such as "https" in URLs) from the content. Finally, using the DBpedia Spotlight[7] Named Entity Recognition (NER) tool, entities were extracted from the pre-processed text to be recommended to the user [13].

A sample recommendation process is shown by the red dashed rectangles in Figure 1. In this case, a recommendation was triggered when a web page was added to the GNN context. The URL and title of the web page were pre-processed and analyzed by the NER module, resulting in three entities being extracted for recommendation to the user [13].

In the original scenario, all entities recognized from the pre-processed text extracted from context-informative events were directly recommended to the user without further filtering (gray dashed arrow in Figure 1). In contrast, our proposed method introduces an ARP module (shown in green color in Figure 1) to refine the recommendation process by incorporating user feedback on previously recommended entities. The core idea behind ARP is to recommend only those entities that are semantically more closely aligned with the ones the user has already confirmed as relevant. By integrating user feedback into the ER process, we demonstrate a noticeable improvement in the system's performance. The following subsections provide a detailed description of our proposed method: we first explain how entities are semantically represented in our approach using RDF2Vec embeddings, then conduct a comparative analysis of different similarity metrics to determine the most effective one, followed by detailing our semantic similarity-based algorithm, which utilizes the selected metric to enhance ER on the RLKWiC dataset. Finally, we present the results to illustrate how our method outperforms the original recommendation strategy.

4.1. DBpedia RDF2Vec Graph Embeddings

DBpedia, the foundational knowledge base used for representing entities in RLKWiC, is a prominent linked open data resource that structures its knowledge as a Resource Description Framework (RDF) graph. The graph-based nature of RDF makes it inherently complex and challenging to manage using traditional data mining and machine learning techniques. To address this challenge, RDF2Vec was introduced as a method for creating semantic representations of entities in RDF graphs by learning latent numerical vector embeddings [29]. RDF2Vec leverages language modeling techniques traditionally used for word embeddings to generate feature vectors from graph substructures, such as those obtained through Weisfeiler-Lehman Subtree RDF Graph Kernels [30, 31] and random graph walks. These embeddings effectively capture the semantics of entities in a way that is suitable for data mining and predictive modeling tasks [29].

In our study, we utilized the DBpedia RDF2Vec graph embeddings dataset[8] [32] to semantically represent RLKWiC entities to enhance the ER accuracy by enabling the system to recommend entities based on their semantic proximity within the DBpedia KG.

Our method is based on the assumption that entities representing a user context, or at least those relevant to it, should be positioned closer to each other in the embedding space. Consequently, by filtering out recognized entities that are not sufficiently close to previously successful recommendations (entities scored as 1: relevant or 2: representative by the user), we can enhance recommendation precision. To evaluate this assumption, we analyzed the pairwise similarities between the embeddings of entities within each context. For an intuitive understanding, we began by visualizing the similarities.

---

[7]https://www.dbpedia-spotlight.org
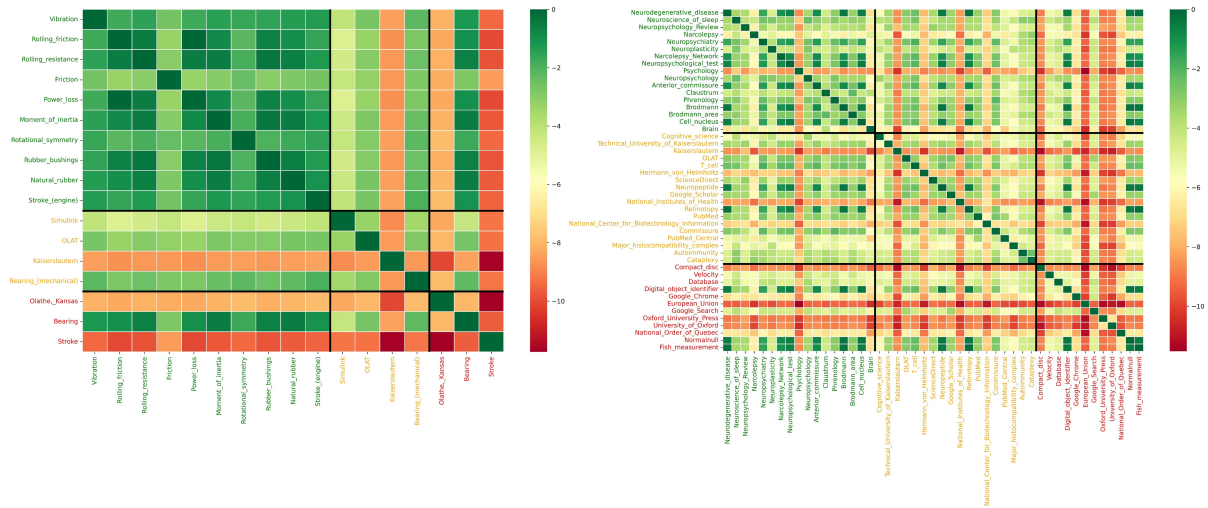[8]Available on https://zenodo.org/records/6377944

Figure 2: Pairwise semantic similarity of entities within two sample contexts (left: "Vehicle Vibrations", and right: "Neuropsychology") from RLKWiC using Euclidean distance.

While similarities based on certain metrics, such as Sigmoid, Linear, and Polynomial kernels, yielded poor results, others aligned well with our hypothesis.

Figure 2 illustrates the pairwise semantic similarity of entities within two sample contexts from RLKWiC using Euclidean distance. The heatmap on the left represents a context named "Vehicle Vibrations" which includes 17 evaluated entities: 10 representative (scored 2, shown in green font), 4 relevant (scored 1, in yellow), and 3 irrelevant (scored 0, in red). The heatmap demonstrates a high pairwise similarity among representative entities, with relevant entities being somewhat less close, and irrelevant ones further away.

An interesting case to note in this context is the entity "stroke". Here, the entity stroke[9] (medical condition) was incorrectly recognized and recommended within the "Vehicle Vibrations" context, whereas the correct entity to recommend was stroke in an engine[10]. The heatmap shows that the correct entity (stroke in engine) is closer to the appropriate entities compared to the incorrect one (stroke as a medical condition), highlighting the potential of semantic information for disambiguation and improving ER. The heatmap on the right depicts another context from a different participant named "Neuropsychology", with more entities, supporting the same interpretation.

While these visualizations provided some intuitive confirmation of our assumption, they are not sufficient for drawing definitive conclusions. Therefore, in the designed experiment presented in the next section, we assess our assumption more rigorously and identify which similarity metrics appear to be more effective in representing entity similarities in alignment with users' assessments of relevance.

4.2. Comparative analysis of similarity metrics

While there has been intriguing research on entity similarities within KG embedding spaces, such as conducting extensive experiments to evaluate the clustering capabilities of various KG embedding models to investigate how different models may capture distinct notions of similarity [33], the challenge of ensuring that these embeddings inherently reflect the user's personal interpretation of entity similarity by positioning similar entities close to each other has not yet been explored. This research takes a step towards this goal by exploring entity similarity in the RDF2Vec DBpedia KG embedding space from the perspective of the user's context. To achieve this, we utilized the scikit-learn library[11], which provides functions for computing pairwise similarities and supports a variety of metrics including Euclidean, Cosine, and Manhattan distances, along with RBF, Laplacian, Sigmoid, Linear, and Polynomial kernels.

---

[9]http://dbpedia.org/resource/Stroke
[10]http://dbpedia.org/resource/Stroke_(engine)
[11]https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise_distances.html

Figure 3: Values and averages of G_0, G_1, and G_2 based on cosine similarity.

To evaluate our assumption and determine which similarity metrics are more effective in aligning with users' assessments of entity relevance to their contexts, we applied a one-way ANOVA (Analysis of Variance) test. The one-way ANOVA is used to determine whether an independent variable has an effect on a dependent variable. In our case, the independent variable is the entity's relevance to the context, which is categorized as 0 (irrelevant), 1 (relevant), or 2 (representative). The dependent variable is the average pairwise semantic similarity between all entities and representative entities within each context. The one-way ANOVA test produces an F-statistic, which is then used to calculate the p-value. The p-value helps us decide whether to reject or fail to reject the null hypothesis (H_0), which states that there is no statistically significant difference between the pairwise semantic similarity within the representative entities of a context compared to their similarity to other entities within the same context. Conversely, the alternative hypothesis (H_1) posits that representative entities are semantically closer to each other. We denote the average pairwise semantic similarities for each group by G_i:

$$G_i = \left\{ \frac{1}{|E_c^2| \cdot |E_c^i|} \sum_{e_2 \in E_c^2} \sum_{e_i \in E_c^i} \text{sim}_m(e_2, e_i) \,\middle|\, c = 1, \ldots, 56 \right\}, \quad \text{for } i = 0, 1, 2$$

Where G_i ($i \in \{1, 2, 3\}$) denotes the set of average semantic similarities of entities scored with i from representative entities (scored with 2) over the 56 existing contexts in RLKWiC. The notation E_c^2 represents the set of entities scored with 2 in context c ($c = 1, \ldots, 56$), and E_c^i represents the set of entities scored with i (which can be 0, 1, or 2) in context c. The function $\text{sim}_m(e_2, e_i)$ denotes the semantic similarity between a representative entity e_2 in E_c^2 and an entity e_i in E_c^i, based on the similarity metric m. The double summation $\sum_{e_2 \in E_c^2} \sum_{e_i \in E_c^i} \text{sim}_m(e_2, e_i)$ calculates the total sum of semantic similarities between all pairs of entities from sets E_c^2 and E_c^i, and dividing by $|E_c^2| \cdot |E_c^i|$ computes the average of these similarities over the number of pairwise comparisons.

Figure 3 presents the values of the three groups calculated using cosine similarity as an example, along with their respective average lines which aligns with our expectation that the similarities among representative entities will be the highest, followed by relevant entities, and then irrelevant ones.

Using the computed average similarity groups across contexts for each metric, we then calculated the F-statistic, which is the ratio of the mean square between groups to the mean square within groups (MST/MSE), to determine the p-values for testing our hypothesis. As shown in Table 1, most p-values are below 0.05, indicating a statistically significant difference between the pairwise semantic similarity among representative entities in a context and their similarity to other entities within the same context.

Table 1
Comparative analysis of similarity metrics using one-way ANOVA (rounded to three decimal places).

	Cosine similarity	Euclidean distance	Manhattan distance	RBF kernel	Laplacian kernel	Sigmoid kernel	Polynomial kernel	Linear kernel
F-statistic	3.503	<u>4.775</u>	4.657	3.961	**4.930**	1.778	2.064	1.902
P-value	0.033	<u>0.010</u>	0.011	0.021	**0.009**	0.173	0.131	0.153

Additionally, the Laplacian kernel and Euclidean distance metrics show the greatest potential for aligning with users' personal interpretations of entity similarities in relation to their contexts.

4.3. Recommendation algorithm

Based on the findings from our semantic analysis, we propose an algorithm aimed at enhancing entity recommendations on the RLKWiC dataset. In our approach, detailed in Algorithm 1, we categorize evaluated entities into relevant (scored 1 or 2) and irrelevant (scored 0) based on the explicit feedback provided in the RLKWiC benchmark dataset. For each newly recognized entity, we compute its semantic distance using the Laplacian kernel metric relative to both the relevant and irrelevant groups. The new entity is recommended only if it is closer to the relevant entities than to the irrelevant ones. We ordered the recognized entities within each context chronologically, simulating the availability of user feedback at each recommendation point, and then applied our recommendation method across the 56 contexts of the 8 participants in the RLKWiC dataset.

Algorithm 1: Adaptive relevance prediction via semantic similarity of RDF2Vec entity embeddings

---

Input: Participants P, for each $p \in P$ a set of contexts C_p including a chronologically ordered set of entities E_{pc}, Participant's explicit feedback on previous recommendations divided into relevant entities R_{pc} and irrelevant entities I_{pc}, scaling parameter $\gamma = \frac{1.0}{200}$

(RDF2Vec embeddings dataset consists of 200-dimensional vectors of DBpedia entities)

Output: Recommendation decision for all recognized entities in RLKWiC

for *each participant* $p \in P$ **do**

    for *each context* $c \in C_p$ **do**

        for *each entity* $e \in E_{pc}$ **do**

            Compute the average semantic similarity of e with relevant entities in C_p:

            $\text{sim}_{rel_{pc}} \leftarrow \frac{1}{|R_{pc}|} \sum_{r \in R_{pc}} \exp\left(-\frac{\|\text{RDF2Vec}(e) - \text{RDF2Vec}(r)\|_1}{\gamma}\right)$

            Compute the average semantic similarity of e with irrelevant entities in C_p:

            $\text{sim}_{irr_{pc}} \leftarrow \frac{1}{|I_{pc}|} \sum_{i \in I_{pc}} \exp\left(-\frac{\|\text{RDF2Vec}(e) - \text{RDF2Vec}(i)\|_1}{\gamma}\right)$

            if $sim_{rel_{pc}} > sim_{irr_{pc}}$ **then**

                Recommend the entity e

                Add e to relevant entities R_{pc}

            else

                Do not recommend the entity e

                Add e to irrelevant entities I_{pc}

The step by step process of simulating the proposed ER method on RLKWiC is explained as follows:

4.3.1. Input Data:

The algorithm works with data from multiple participants. For each participant, the algorithm processes a set of contexts, each of which includes a chronologically ordered set of entities. The participant's explicit feedback is also provided, where entities are divided into two categories:

- *Relevant entities* (scored 1 or 2), meaning the user found these entities useful.
- *Irrelevant entities* (scored 0), meaning the user did not find these entities useful.

The core data utilized for the similarity computation is a set of 200-dimensional RDF2Vec embeddings representing DBpedia entities. These embeddings capture semantic information about the entities.

4.3.2. Contextual Evaluation:

The algorithm processes the entity data within each participant's context. For each context, entities are evaluated one by one in chronological order, simulating a scenario where the system provides recommendations progressively, as more feedback becomes available.

4.3.3. Similarity Computation:

For each entity, the algorithm calculates two separate semantic similarity scores:

- *Similarity to relevant entities*: This measures how close the new entity is to entities previously marked as relevant within the current context. The similarity is computed using the Laplacian kernel, which measures the distance between the RDF2Vec embedding of the new entity and the embeddings of the relevant entities. The formula for the similarity is:

$$\text{sim}_{rel_{pc}} = \frac{1}{|R_{pc}|} \sum_{r \in R_{pc}} \exp\left(-\frac{\|\text{RDF2Vec}(e) - \text{RDF2Vec}(r)\|_1}{\gamma}\right)$$

  where $\|\cdot\|_1$ denotes the L1 (Manhattan) distance, and γ is a scaling parameter that controls the sensitivity of the similarity function.
- *Similarity to irrelevant entities*: This is similarly computed, but compares the new entity to those previously marked as irrelevant, using the same Laplacian kernel approach:

$$\text{sim}_{irr_{pc}} = \frac{1}{|I_{pc}|} \sum_{i \in I_{pc}} \exp\left(-\frac{\|\text{RDF2Vec}(e) - \text{RDF2Vec}(i)\|_1}{\gamma}\right)$$

4.3.4. Recommendation Decision:

Once both similarity scores are computed, the algorithm compares them:

- If the similarity to relevant entities ($\text{sim}_{rel_{pc}}$) is greater than the similarity to irrelevant entities ($\text{sim}_{irr_{pc}}$), the new entity is recommended. This means the new entity is considered closer, in terms of semantic similarity, to the group of relevant entities.
- If the similarity to irrelevant entities is higher, the entity is not recommended.

4.3.5. Feedback Incorporation:

After each decision (whether to recommend or not), the algorithm updates the relevant or irrelevant entity sets for that context by adding the newly evaluated entity to the corresponding set. This allows the recommendation process to evolve as more user feedback becomes available.

Table 2

Performance comparison between the baseline method and ER with ARP on RLKWiC.

Recommendation task	Method	Accuracy	Precision	Recall	F1-Score
Recommending relevant entities	Baseline ER	0.563	0.563	**1.00**	**0.721**
	ER with ARP	**0.598**	**0.613**	0.780	0.686
Recommending representative entities	Baseline ER	0.258	0.258	**1.00**	0.410
	ER with ARP	**0.458**	**0.302**	0.838	**0.444**

4.3.6. Repeat for All Participants and Contexts:

The process is repeated for each participant and for each context within their data to generate a set of recommendation decisions for all recognized entities in the RLKWiC dataset. These recommendations are personalized for each participant and context, based on how semantically similar a new entity is to those previously marked as relevant or irrelevant.

In summary, this algorithm adapts to user preferences over time by learning from their explicit feedback and progressively improving recommendations based on the semantic similarities of entity embeddings. The use of the Laplacian kernel with RDF2Vec embeddings allows for the assessment of how closely related new entities are to those the user has previously interacted with, ultimately aiming to deliver more relevant recommendations.

4.4. Results

In our evaluation, we focused exclusively on the set of entities recognized by the NER module within the scenario, excluding any additional entities manually added by users (Unlike the original evaluation [13]). Since the baseline method recommends all entities identified by the NER, it achieves the highest possible recall value (1.0). In contrast, our ARP-based method filters out entities that are not semantically related to the relevant ones, resulting in improved accuracy. The results presented in Table 2 provide a detailed comparison of the performance of the two methods. While ARP only provides a slight improvement in recommending relevant entities, it enhances the accuracy for representative entities by 20 percent.

Additionally, we trained several binary classification models, including Random Forest, Linear Regression, and Gaussian Naive Bayes, which are known to perform well with sparse datasets, to predict the relevance of recognized entities using their embeddings. However, these models did not demonstrate any significant improvements compared to our proposed ARP-based approach.

5. Conclusion and Future Work

In this paper, we explored ER as a method to support knowledge work. We examined the evolution of ER tools towards PKA and introduced the RLKWiC benchmark, the only existing dataset that ensures full transparency and reproducibility for ER evaluation. Our approach outperformed the defined baseline by 20% in recommending representative entities by incorporating user feedback and simulating ER using an ARP module.

To predict the relevance of recognized entities in an adaptive manner, we measured the semantic distance between new entities and previously evaluated entities within each context. We utilized RDF2Vec embeddings to represent DBpedia entities in the RLKWiC dataset. Additionally, we tested our core hypothesis statistically, demonstrating the potential of semantic-based methods in enhancing ER. We found that among various similarity metrics, Laplacian kernel and Euclidean distance show more potential in maintaining the user's context-based interpretation of entity closeness.

Furthermore, We highlighted that entities perceived as similar by a user may not align with their closeness in conventional knowledge representations. Consequently, methods that are effective in general knowledge tasks may not necessarily perform well in PKA, where a more personalized

interpretation of world knowledge is required. Further exploration of novel approaches to computing semantic similarity of resources in DBpedia [34] could offer valuable insights in this regard.

Despite the usefulness of the RLKWiC dataset for evaluating our approach, the study has some limitations. The small sample size, with only eight participants, restricts the generalizability of our findings, making it essential to explore the robustness of the results with a larger and more diverse user base in future work. Additionally, our study could benefit from a more comprehensive comparison with alternative ER approaches. Benchmarking our approach against state-of-the-art ER methods would provide deeper insights into its relative performance and areas for further improvement. Addressing these limitations in future research could strengthen the validity and applicability of our conclusions.

Further avenues could be pursued to further improve ER. One potential enhancement involves incorporating a disambiguation module into the system. This would allow the model to distinguish context-specific meanings of entities, such as recognizing that "stroke" in the context of "Vehicle Vibrations" refers to an engine component rather than a medical condition.

Our method focuses on enhancing the ER scenario following the NER phase. However, this enhancement could also be applied at an earlier stage in the ER process by leveraging novel methods, such as inflection-tolerant ontology-based NER for real-time applications [35], to further improve the effectiveness of ER.

Additionally, large language models are anticipated to excel in predicting entity relevance when given enough information about user contexts, as they have already demonstrated considerable success in various domains including RS[36].

6. Acknowledgments

References

[1] H. Ko, S. Lee, Y. Park, A. Choi, A survey of recommendation systems: Recommendation models, techniques, and application fields, Electronics 11 (2022). URL: https://www.mdpi.com/2079-9292/11/1/141.

[2] J. B. Schafer, D. Frankowski, J. Herlocker, S. Sen, Collaborative filtering recommender systems, in: The adaptive web, Springer, 2007, pp. 291–324.

[3] V. W. Anelli, P. Basile, G. De Melo, F. M. Donini, A. Ferrara, C. Musto, F. Narducci, A. Ragone, M. Zanker, Fifth knowledge-aware and conversational recommender systems workshop (KaRS), in: Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 1259–1262. doi:10.1145/3604915.3608759.

[4] V. W. Anelli, P. Basile, D. Bridge, T. Di Noia, P. Lops, C. Musto, F. Narducci, M. Zanker, Knowledge-aware and conversational recommender systems, RecSys '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 521–522. doi:10.1145/3240323.3240338.

[5] M. Bakhshizadeh, C. Jilek, H. Maus, A. Dengel, Leveraging context-aware recommender systems for improving personal knowledge assistants by introducing contextual states, LWDA, 2021, pp. 1–12. URL: https://ceur-ws.org/Vol-2993/paper-01.pdf.

[6] T. H. Davenport, Thinking for a Living: How to Get Better Performances and Results from Knowledge Workers, Harvard Bus. School Press, 2005.

[7] P. F. Drucker, Knowledge-worker productivity: The biggest challenge, California Management Review 41 (1999) 79–94.

[8] M. Bakhshizadeh, C. Jilek, H. Maus, A. Dengel, Towards context-aware recommender systems for supporting knowledge workers in personal and corporate information space, 2024.

[9] O. Rostanin, H. Maus, T. Suzuki, K. Maeda, Using concept maps to improve proactive information delivery in tasknavigator, in: R. Setchi, I. Jordanov, R. J. Howlett, L. C. Jain (Eds.), Knowledge-Based and Intelligent Information and Engineering Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 639–648. doi:10.1007/978-3-642-15387-7_67.

[10] M. Sappelli, S. Verberne, W. Kraaij, Evaluation of context-aware recommendation systems for information re-finding, Journal of the Association for Information Science and Technology 68 (2017) 895–910. doi:10.1002/ASI.23717.

[11] T. Vuong, S. Andolina, G. Jacucci, P. Daee, K. Klouche, M. Sjöberg, T. Ruotsalo, S. Kaski, Entitybot: Actionable entity recommendations for everyday digital task, in: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, CHI EA '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3491101.3519910.

[12] T. Vuong, S. Andolina, G. Jacucci, P. Daee, K. Klouche, M. Sjöberg, T. Ruotsalo, S. Kaski, Entitybot: Supporting everyday digital tasks with entity recommendations, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 753–756. doi:10.1145/3460231.3478883.

[13] M. Bakhshizadeh, H. Maus, A. Dengel, Context-based entity recommendation for knowledge workers: Establishing a benchmark on real-life data, in: Proceedings of the 18th ACM Conference on Recommender Systems, RecSys '24, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3640457.3688068.

[14] Z. Wang, F. Wang, J.-R. Wen, Z. Li, Bring user interest to related entity recommendation, in: M. Croitoru, P. Marquis, S. Rudolph, G. Stapleton (Eds.), Graph Structures for Knowledge Representation and Reasoning, Springer International Publishing, Cham, 2015, pp. 139–153.

[15] X. Yu, X. Ren, Y. Sun, Q. Gu, B. Sturt, U. Khandelwal, B. Norick, J. Han, Personalized entity recommendation: a heterogeneous information network approach, in: Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM '14, Association for Computing Machinery, New York, NY, USA, 2014, p. 283–292. doi:10.1145/2556195.2556259.

[16] E. Palumbo, D. Monti, G. Rizzo, R. Troncy, E. Baralis, entity2rec: Property-specific knowledge graph embeddings for item recommendation, Expert Systems with Applications 151 (2020) 113235. doi:/10.1016/j.eswa.2020.113235.

[17] F. M. Harper, J. A. Konstan, The movielens datasets: History and context, ACM Trans. Interact. Intell. Syst. 5 (2015). doi:10.1145/2827872.

[18] T. Vuong, G. Jacucci, T. Ruotsalo, Proactive information retrieval via screen surveillance, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1313–1316. doi:10.1145/3077136.3084151.

[19] T. Vuong, G. Jacucci, T. Ruotsalo, Watching inside the screen: Digital activity monitoring for task recognition and proactive information retrieval, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (2017). doi:10.1145/3130974.

[20] C. Lampasona, O. Rostanin, H. Maus, Seamless integration of order processing in ms outlook using smartoffice: an empirical evaluation, in: Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM '12, Association for Computing Machinery, New York, NY, USA, 2012, p. 165–168. doi:10.1145/2372251.2372281.

[21] T. Vuong, M. Saastamoinen, G. Jacucci, T. Ruotsalo, Understanding user behavior in naturalistic information search tasks, Journal of the Association for Information Science and Technology 70 (2019) 1248–1261. doi:/10.1002/asi.24201.

[22] G. Jacucci, P. Daee, T. Vuong, S. Andolina, K. Klouche, M. Sjöberg, T. Ruotsalo, S. Kaski, Entity recommendation for everyday digital tasks, ACM Trans. Comput.-Hum. Interact. 28 (2021). doi:10.1145/3458919.

[23] T. Vuong, S. Andolina, G. Jacucci, T. Ruotsalo, Does more context help? effects of context window and application source on retrieval performance, ACM Trans. Inf. Syst. 40 (2021). doi:10.1145/3474055.

[24] T. Vuong, Behavioral Task Modeling for Entity Recommendation, Ph.D. thesis, Finland, 2022.

[25] Z. R. Yousefi, T. Vuong, M. AlGhossein, T. Ruotsalo, G. Jaccuci, S. Kaski, Entity footprinting: Modeling contextual user states via digital activity monitoring, ACM Trans. Interact. Intell. Syst. 14 (2024). doi:10.1145/3643893.

[26] T. Vuong, T. Ruotsalo, Predicting representations of information needs from digital activity context, ACM Trans. Inf. Syst. 42 (2024). doi:10.1145/3639819.

[27] M. Bakhshizadeh, Supporting knowledge workers through personal information assistance with context-aware recommender systems, in: Proceedings of the 18th ACM Conference on Recommender Systems, RecSys '24, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3640457.3688010.

[28] M. Bakhshizadeh, C. Jilek, M. Schröder, H. Maus, A. Dengel, Data collection of real-life knowledge work in context: The rlkwic dataset, in: S. Li (Ed.), Information Management, Springer Nature Switzerland, Cham, 2024, pp. 277–290. doi:10.1007/978-3-031-64359-0_22.

[29] P. Ristoski, H. Paulheim, Rdf2vec: Rdf graph embeddings for data mining, in: P. Groth, E. Simperl, A. Gray, M. Sabou, M. Krötzsch, F. Lecue, F. Flöck, Y. Gil (Eds.), The Semantic Web – ISWC 2016, Springer International Publishing, Cham, 2016, pp. 498–514.

[30] G. K. D. de Vries, A fast approximation of the weisfeiler-lehman graph kernel for rdf data, in: H. Blockeel, K. Kersting, S. Nijssen, F. Železný (Eds.), Machine Learning and Knowledge Discovery in Databases, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 606–621.

[31] G. K. D. de Vries, S. de Rooij, Substructure counting graph kernels for machine learning from rdf data, Journal of Web Semantics 35 (2015) 71–84. doi:10.1016/j.websem.2015.08.002, machine Learning and Data Mining for the Semantic Web (MLDMSW).

[32] M. P. Christensen, M. Lissandrini, K. Hose, Dbpedia rdf2vec graph embeddings, 2022. doi:10.5281/zenodo.6377944.

[33] N. Hubert, H. Paulheim, A. Brun, D. Monticolo, Do similar entities have similar embeddings?, in: A. Meroño Peñuela, A. Dimou, R. Troncy, O. Hartig, M. Acosta, M. Alam, H. Paulheim, P. Lisena (Eds.), The Semantic Web, Springer Nature Switzerland, Cham, 2024, pp. 3–21.

[34] G. Piao, S. s. Ara, J. G. Breslin, Computing the semantic similarity of resources in dbpedia for recommendation purposes, in: G. Qi, K. Kozaki, J. Z. Pan, S. Yu (Eds.), Semantic Technology, Springer International Publishing, Cham, 2016, pp. 185–200.

[35] C. Jilek, M. Schröder, R. Novik, S. Schwarz, H. Maus, A. Dengel, Inflection-tolerant ontology-based named entity recognition for real-time applications, in: M. Eskevich, G. de Melo, C. Fäth, J. P. McCrae, P. Buitelaar, C. Chiarcos, B. Klimek, M. Dojchinovski (Eds.), 2nd Conference on Language, Data and Knowledge (LDK 2019), volume 70 of *Open Access Series in Informatics (OASIcs)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2019, pp. 11:1–11:14. doi:10.4230/OASIcs.LDK.2019.11.

[36] L. Wu, Z. Zheng, Z. Qiu, H. Wang, H. Gu, T. Shen, C. Qin, C. Zhu, H. Zhu, Q. Liu, H. Xiong, E. Chen, A survey on large language models for recommendation, World Wide Web 27 (2024) 60. doi:10.1007/s11280-024-01291-2.