Evaluating Explainable Interfaces for a Knowledge Graph-Based Recommender System

Erasmo Purificato[1,2], Baalakrishnan Aiyer Manikandan[1], Prasanth Vaidya Karanam[1], Mahantesh Vishvanath Pattadkal[1] and Ernesto William De Luca[1,2]

[1]Otto von Guericke University Magdeburg, Universitätsplatz, 2, 39106 Magdeburg, Germany
[2]Georg Eckert Institute for International Textbook Research, Celler Straße, 3, 38114 Brunswick, Germany

Abstract

In this paper, we present the design and the implementation of a knowledge graph-based recommender system for research paper suggestion, along with two explainable interfaces which provide different types of explanations to the users interacting with the recommender. Our work, developed within the academic context of the Georg Eckert Institute for International Textbook Research, aims to assess the effectiveness of the explanation among the researchers of the institute and understand which characteristics of the interfaces themselves are perceived to be as most interpretable, leading to increase the trust and confidence in the recommender system and its credibility. We evaluated our work through a user study performed among different experts covering several research fields. All participants were asked to take part in an online survey, and a focus group answered some targeted interviews. This last qualitative evaluation aims better to understand the interaction patterns within the two explainable interfaces. The results show the greater effectiveness of the interface providing the explanation through a natural language sentence and displaying the graph path from the user to the recommended paper.

Keywords

knowledge graph-based recommender system, explainable systems, explainable interfaces

1. Introduction

Recommenders and user-adaptive systems, as well as search engines, may be considered the most popular technologies of the current digital information age. The advent and fast spread of recommendation systems have contributed significantly to the growth of interests in retrieving personalised information in several contexts, from e-commerce to academic research, for the latter mostly in terms of experts [1, 2] and scientific paper recommendations [3, 4]. It is self-evident to point out that in the era of big data and information overload, having such systems can help in navigating the mass of content being created on a daily basis, especially for academics, for whom not being aware of relevant works, experts or interesting research projects is a common problem.

IntRS'21: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, September 25, 2021, Virtual Event
erasmo.purificato@ovgu.de (E. Purificato); baalakrishnan.aiyer@st.ovgu.de (B. A. Manikandan); prasanth.karanam@st.ovgu.de (P. V. Karanam); mahantesh.pattadkal@st.ovgu.de (M. V. Pattadkal); ernesto.deluca@ovgu.de (E. W. De Luca)
0000-0002-5506-3020 (E. Purificato); 0000-0003-3621-4118 (E. W. De Luca)

Hand in hand with the existing desire for content customisation, we are now also facing the increasing need for interpretability of the system outcomes from the end-user perspective. Explainable artificial intelligence (XAI) is a key research area in computer science having the goal of "*exposing complex artificial intelligence models to humans in a systematic and interpretable manner*" [5], and thus dealing with the creation of transparent, human-understandable and trustworthy intelligent systems.

In such a scenario, the design of user interfaces (UIs) also plays a fundamental role to provide the proper explanations to the end-users, even more than the implementation of the system itself in many cases, and even though a lot of exciting works have been published about explainability methods over the past few years (e.g. [6, 7]), the notion of "right" or "good" explanation is steadily under study in the user-centric research area. Some works that inspired this paper show that different goals and cognitive capabilities affect the perception of explanation [8] and different users require different explanation details [9], while at the same time different individual characteristics can even change the perception of transparency [7].

In this paper, we present a knowledge graph-based recommender system (also referred to in this work simply as graph-based recommender) for scientific publications combined with two explainable interfaces designed to provide different explanations to the users interacting with the recommender. We aim to assess which of the different ways of displaying results and related explanations is the most comprehensible and likely to increase confidence and trustworthiness in using the system by researchers covering different fields of study, such as computer science, educational media, linguistics, social sciences and humanities.

Our work is developed in the academic context of the Georg Eckert Institute for International Textbook Research[1] (hereafter "GEI" or "the institute"). The GEI, member of the Leibniz Association and located in Brunswick (Germany), conducts international, multidisciplinary and application-oriented research into school textbooks and educational media, centring on approaches drawn from historical and cultural studies.

The GEI uses the Elsevier's Research Information Management System *Pure* [2], which provides a structured, relational data model that links together all content types within the system and allows both for a full view of the institution's research activity and output and detailed reporting across the research lifecycle. Furthermore, the researchers are asked to update their own data. These data are stored in a user profile, including their department affiliation, the job information, published works, research interests, projects in which they are involved and existing relationships with externals (e.g. co-authorship). They are used to build up the recommendations.

In this work, we implemented our recommender system starting from the idea given by *entity2rec* [10], a technique presented in 2017 by Palumbo et al. to measure user-item relatedness for top-N item recommendation. This methodology leads to the definition of several subgraphs from the original knowledge graph, considering one property (i.e. a relationship between two entities) at a time. As described in detail in the continuation of this paper, each of these *property-specific subgraphs* is used to compute the similarity between users and papers.

The two explainable interfaces, which hereafter will be referred to as *System A* and *System B*,

[1] http://www.gei.de/en/home.html. Last seen August 27, 2021.
[2] https://www.elsevier.com/solutions/pure. Last seen August 27, 2021.

are designed, respectively, with the following concepts: System A provides first a *one-line explanation*, expressing the main contribution for the specific recommendation (e.g. "*Recommendation based on your research interests and activities*"), and then displays the *path* on the knowledge graph from the user and the suggested paper; on the other hand, System B shows a *percentage score*, indicating the average similarity score of all the contributions to the recommendation. Then the detailed explanation is provided through a *bar chart* displaying the individual similarity score for each property.

The evaluation of the graph-based recommender and the explainable interfaces is carried out by means of a user study conducted among the researchers at the GEI, through a guided tour of the UIs, in order to get to know and become familiar with them before answering an online survey for the actual assessment. The user study aims to assess the effectiveness of both explainable interfaces, and a qualitative evaluation is performed through a *5-point Likert scale*.

Finally, to better understand the interaction patterns within the two explainable interfaces, some targeted interviews are conducted.

The remainder of the paper is structured as follows: in Section 2 we present the related work of the last decade about graph-based recommender systems and the explainability of their recommendations; the used knowledge graph is described in Section 3; the design and implementation of the presented recommender system and explainable interfaces are discussed in Section 4; the empirical evaluation of our work with the related results are illustrated in Section 5; finally, in Section 6 conclusion and future works are discussed.

2. Related work

Among the various definitions of *knowledge graph* (KG) that have been surveyed over the years, the most inclusive one has been provided by Hogan et al. [11] and describe a knowledge graph as "*a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities*". KG-based recommenders are currently a fervent research topic, and this is mainly due to particular characteristics of these systems: mapping items and their attributes can discover mutual relations between items into the KG [12]; user and user-side information can also be integrated into the KG, making it possible to detect user-item relations and user preferences in an efficient and accurate way [13]; the results of a KG-based recommender are easily interpretable [14]. Recently, Guo et al [15] categorised the KG-based recommender systems in three ways, depending on how they apply KGs: *path-based* methods, *embedding-based* methods and *unified* methods.

Path-based methods create a user-item graph and exploit the connectivity patterns of entities in the KG for providing recommendations. Generally, these models take advantage of the connectivity similarity (i.e. user-user, item-item and user-item) to enhance the recommendations. In most cases, they employ semantic similarities of entities in different meta-paths as regularisation for refining the latent vector representations of users and items [16]. One of the main disadvantages of this kind of models is the high dependence of the number and type of meta-paths to the specific application domain; thus, their use is not easily exportable to any data. On the other hand, path-based models inherently keep interpretability in their recommendations by the similarity between users and items on the meta-path level. For example, the meta-path

"$user \rightarrow item \rightarrow user \rightarrow item$" is used for collaborative recommendations, and this can also be translated into a natural language explanation just following the connections.

Embedding-based models generate a dense low-rank vector representation of entities and the relations between them, which can later be compared using similarity measures to model user preferences and yield item recommendations [17]. Some common applications, like *entity2rec* [10], leverage graph-embedding algorithms, such as *node2vec* [18], which makes use of the concept of *random walks* to sample sequences of nodes to be treated as "words" in a document, and then learning the node embeddings by using neural language models, such as *word2vec* [19]. This approach takes into account the unique connectivity patterns between the nodes of a network, interpreting them as closed-knit communities called *homophily*, which is a representation of their contextual similarity in the network. One of the advantages of an embedding-based model is what Herlocker et al. [20] defines as *serendipity* and *novelty*, namely the ability to recommend items that are new, original, even unusual and unexpected for the user, but still relevant to him/her. While path-based models are constrained by the existing edges between the entities in KG, embedding-based models explore item candidates in feature spaces where KG edge constraints do not exist [21]. Embeddings generally are not considered very explainable because they are low-rank sub-symbolic representations of KGs. Few models in combination with embedding learn first-order logic rules, which can be used as explanations in these models. These models are constrained by the domain-specific interpretation of the graph, entities, and their relations [17].

The unified method benefit from the advantages of both semantic graph-embeddings and semantic path patterns. These methods exploit the idea of embedding propagation to refine the representation of the item or user with multi-hop neighbours in the KG [15] and generally adopt a *graph neural network* architecture to the scope. As easily understandable, these methods inherit interpretability from path-based models.

3. GEI Knowledge Graph

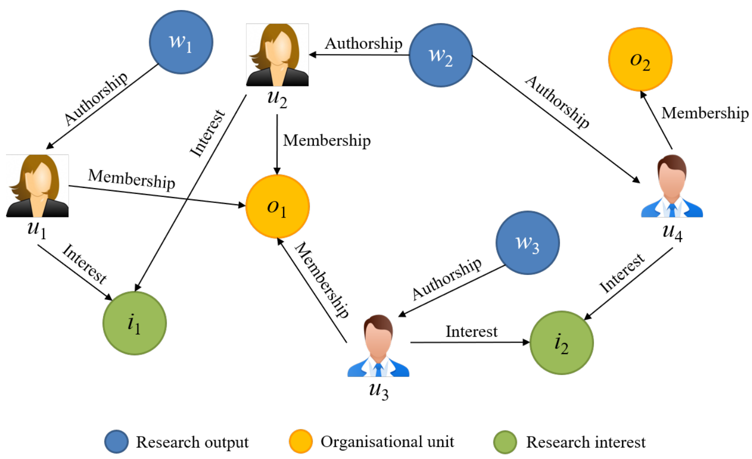


Figure 1: Illustrative representation of the used knowledge graph.

The knowledge graph used in this work is mainly derived from the Pure management system instance of the institute. Along with GEI members' profiles, it contains the users' job information, published works, research interests, projects in which they are involved and existing relationships with external persons (e.g. co-authorship). Data related to persons, projects, research outputs, organisational units and every connection between entities is retrieved by using the Pure Web Service (v5.18[3]) and stored in a Neo4j[4] graph database. The current version of the GEI KG is composed by 6921 nodes and 8988 relationships, whose distribution are respectively displayed in Table 1 and Table 2.

Table 1
Entity types within the GEI Knowledge Graph. For each entity, the short notations are within brackets.

Entity label	Count	Percentage	Entity label	Count	Percentage
Activity (ac)	1679	24.26%	Person (u)	343	4.96%
Course (c)	110	1.59%	Press-Media (pm)	408	5.9%
Event (ev)	829	11.98%	Prize (pz)	7	0.1%
External organisation (eo)	1359	19.64%	Project (p)	68	0.98%
External person (eu)	559	8.08%	Publisher (pb)	301	4.35%
Journal (j)	239	3.45%	Research interest (i)	23	0.33%
Organisational unit (o)	12	0.17%	Research output (w)	984	14.22%

Table 2
Relationship types within the GEI Knowledge Graph. For each relation, we reported the connection it represents (the notation $\{e_1, e_2\} \rightarrow \{e_3, e_4\}$ means that every entity in the first set can be linked to every entity of the second set through the relationship).

Relationship	Count	Percentage	Relationship	Count	Percentage
Association ($\{ac, pm\} \rightarrow u$)	1966	21.87%	Interest ($u \rightarrow i$)	79	0.88%
Authorship ($w \rightarrow \{u, eu\}$)	1839	20.46%	Journalisation ($a \rightarrow j$)	28	0.31%
Creation ($c \rightarrow u$)	112	1.25%	Management ($\{ac, c, p, pm, pz\} \rightarrow \{o, eo\}$)	2286	25.43%
Having-activity ($\{p, pm\} \rightarrow ac$)	231	2.57%	Membership ($\{u, eu\} \rightarrow \{o, eo\}$)	934	10.39%
Having-event ($ac \rightarrow ev$)	1088	12.11%	Participation ($p \rightarrow u$)	150	1.67%
Having-output ($p \rightarrow w$)	108	1.2%	Publication ($j \rightarrow pb$)	160	1.78%

[3]https://api.research-repository.uwa.edu.au/ws/api/518/api-docs/index.html. Last seen August 27, 2021.
[4]https://neo4j.com/. Last seen August 27, 2021.

4. System implementation

In this section, we discuss the design and implementation of the KG-based recommender system and the two explainable interfaces, which are subdivided in four steps: *data preparation*, *topic creation for research outputs*, *recommender system implementation* and *explainable interfaces design*. Each part is described in a separate section below.

4.1. Data preparation

Given the scope of our work to build a system for paper recommendations, we focus primarily on the *authorship* relation between *research outputs* and *persons* (or *external persons*) as the pivotal connection of the entire system, around which subsequent steps are developed.

- **Research output** entities represent the research papers published by the GEI members, and they are represented in the KG with *title*, *abstract*, *language* and authorship connections. There are 984 research outputs scattered across various languages (53% papers are written in German, 38% in English, 4% in other languages, such as Polish, French, Russian and Italian, and the rest is undefined). The title is available for all the nodes, while the abstract is only for 219 of them.
- **Person** entity represents the employees of the GEI with details about their personal data, as well as their job position. In total, there are 343 employees, with both academic and non-academic staff type. Out of 343, only 92 persons have a research output contribution, and they form the *user group* for our recommendation system.
- **External person** nodes belong to people who have collaborated with researchers at the GEI and produced research outputs that are part of the KG. Although our recommender focuses on only employees of GEI, the external person nodes have also been considered as they are directly linked to the research output nodes. In total, there are 559 external persons, and 368 of them have at least one authorship relation.

Apart from research outputs and other entities, person nodes are also directly linked to *organisational units*, *research interests* and *activities*, which might be crucial in our implementation to generate the *user profiles* through which recommendations can be produced.

- **Organisational unit** entities basically represent the internal *departments* at the GEI. Among all the departments, we only consider in our work those that are connected to at least one person with an authorship relation, and the distribution is shown in Table 3. In the presented system and case study, we did not take into account *external organisational units* and their connections with external persons. It is worth noting that some of the employees belong to two or more departments and that out of the 92 persons with authorship relationships, only 85 of them belong to any of the departments. This suggests that not all employees have membership relationships correctly set, opening room for further discussion and analysis of data stored in the Pure management system.
- **Research interest** nodes indicate the research area of interest to users. Out of the 92 persons with research paper contribution, only 30 of them have a research interest in the KG. In total, there are 23 research interests, including *artificial intelligence*, *machine*

Table 3
Distribution of *person* entities across *departments*, reported with the original German name.

Organisational unit	Persons count
Digitale Informations- und Forschungsinfrastrukturen	21
Europa. Narrative, Bilder, Räume	22
Forschungsbibliothek	8
Mediale Transformationen	13
Schulbuch als Medium	19
Schulbuch und Gesellschaft	15
Wissen im Umbruch	17

learning, *recommender systems*, *usability*, *education*, *history*, *digital humanities*, and others. Although the research interest entity is an essential feature for our recommender, the sparse nature of those nodes about persons does not enable us to use them properly in the system implementation process. Therefore, we decided to merge them with the **activity** nodes, which represent all the events that GEI members participated in, such as guest lectures, conferences, workshops, and their role in it. It is important to note that out of the 92 employees with an authorship relationship, only 74 of them are connected to any of the activities.

In Fig. 1 an illustrative representation of the KG resulting from the data preparation is shown.

4.2. Topic creation for research outputs

In order to further characterise the *research outputs* and thus enhance the KG, we employed a **topic modelling** algorithm on those entities to generate five broad topics which would best fit the data.

For the *corpus* preparation, the title and the abstract (when available) have been merged to form a single document representing a research paper. Since the most common topic modelling algorithms are developed for the English language, each document has been translated to English (if needed) to create a uniform corpus. The algorithm used to derive the distinct topics is *BERTopic* [22]. BERTopic is a topic modelling technique based on *BERT embeddings*. Once the embedding are generated by means of a *sentence-transformer*, they are reduced to low-dimension vectors using the *UMAP* algorithm [23], and then further passed to the *HDBSAN* algorithm [24] to perform density-based clustering. For the visualisation of the topics, a class-based TF-IDF is used whilst keeping important words in the topic descriptions. For all the abovementioned algorithms, we used default parameter values.

Table 4 displays the results of the topic modelling procedure. The employed algorithm has been able to classify 638 research outputs into five distinct topics. After the topic modelling, a new node for each topic is created, added to the KG and linked to the related research paper. The updated representation of the KG after topic modelling is shown in Fig. 2.

Table 4
Topic modelling results using *BERTopic*.

Topic label	Representative terms	Papers count
Information retrieval	information, user, semantic	144
International and German history	german, polish, czech, history	87
GEI	eckert, institution, education	72
Educational and textbook	research, studies, education	285
Social, environment and Politics	moral, prejudice, conflict, threat	50

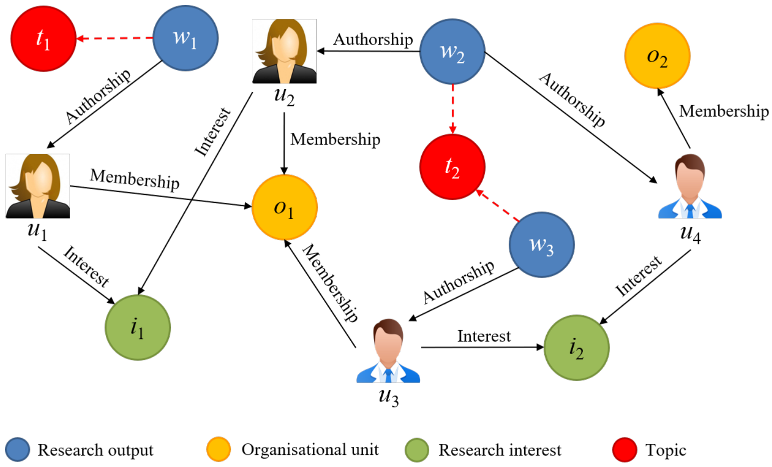


Figure 2: Illustrative representation of the used knowledge graph after topic modelling.

4.3. Recommender system implementation

The presented KG-based recommender system is based on *entity2rec*, an approach to learn user-item (in our case, person-research output) relatedness from KGs for top-N recommendation. Starting from the "whole" KG, for each relationship (called *property*), a property-specific vector representation of entities is learned by exploiting *node2vec* algorithm, and then a property-specific relatedness score is computed using the resulting vector representation.

The property-specific vector representation learning procedure can be seen as the application of the graph embedding algorithm on a *subgraph* extracted from the KG by considering one relationship at a time. Given the *authorship* relationship as pivotal (i.e. always present and thus common to all subgraphs), we generated three subgraphs:

- **Topic subgraph** (Fig. 3), containing *persons*, *research outputs* and associated *topics* generated by the topic modelling algorithm;
- **Research interest and Activity subgraph** (Fig. 4), which includes *research outputs*, *persons*, their *research interests* and the *activities* they were involved in;
- **Department subgraph** (Fig. 5), containing *research outputs*, *persons* and the *organisational units* they belong to.

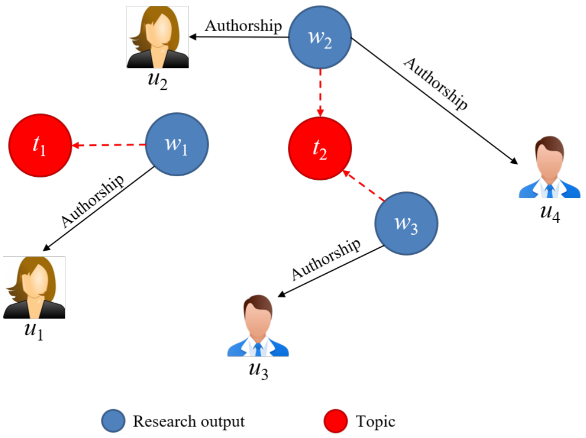


Figure 3: Topic subgraph

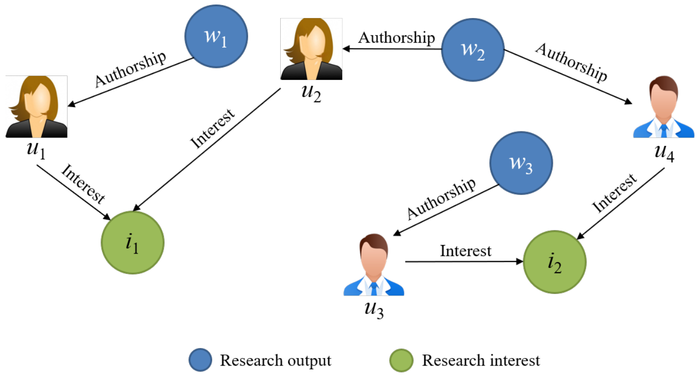


Figure 4: Research interest and Activity subgraph

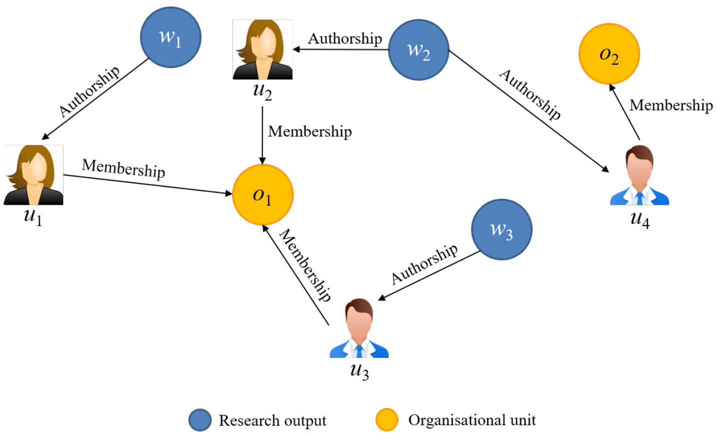


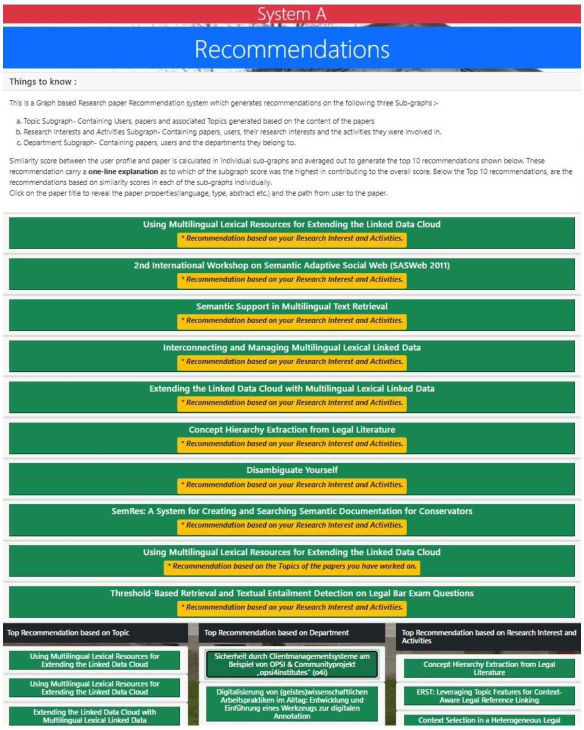
Figure 5: Department subgraph

We aim to produce a list of recommended papers for each of the 92 persons with an authorship relationship in the KG. Using embeddings generated by the node2vec algorithm for each of the subgraphs, we calculate the similarity scores between each person embedding vector and the research output embedding vectors at a subgraph level by using the *cosine similarity* for the practical computation on the vectors. The scores from each of the subgraphs for a person-research output combination are averaged out to obtain the final overall similarity score which will be used to rank our recommendations:

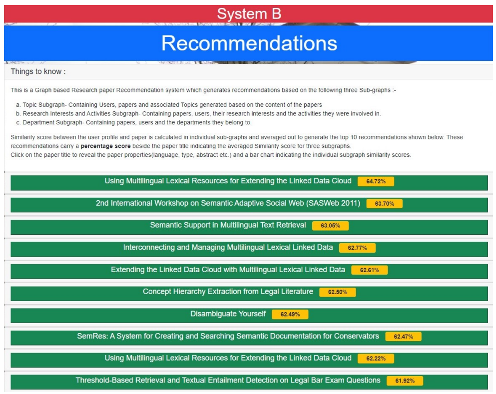
$$sim(u, w) = \frac{1}{|N|} \sum_{i} sim_i(u, w)$$

where *u* and *w* are, respectively, a generic user and a generic research output, N is the number of generated subgraphs and i represents a single subgraph at a time.

4.4. Explainable interfaces design



(a) *System A*



(b) *System B*

Figure 6: Explainable interfaces

The *explainable interfaces* are designed to provide different types of explanations to the users interacting with the system, to assess which of the different ways of displaying the recommendations and the related explanations is the most understandable and likely to instil (or increase) confidence and trustworthiness in the use of the system to researchers covering different fields of study, such as computer science, educational media, linguistics, social sciences and humanities.

The two explainable interfaces, referred to as *System A* and *System B*, are shown in Fig. 6 and respectively implemented as illustrated in the following.

- **System A** is divided in two segments, as displayed in Fig. 6a. The first segment shows the top-10 recommended paper and provides a *one-line explanation*, expressing which subgraph similarity between the user and the research paper is dominant (Fig. 7). Additionally, upon clicking on one recommendation, the user gets the path in the KG traversed by the system to reach that research output node starting from the person node (Fig. 8). The second segment, shown in Fig. 9, presents the top-3 paper recommendation for each individual subgraph.

Figure 7: *System A* - Illustrative portion of the first UI segment displaying 5 out of the top-10 recommended research outputs.

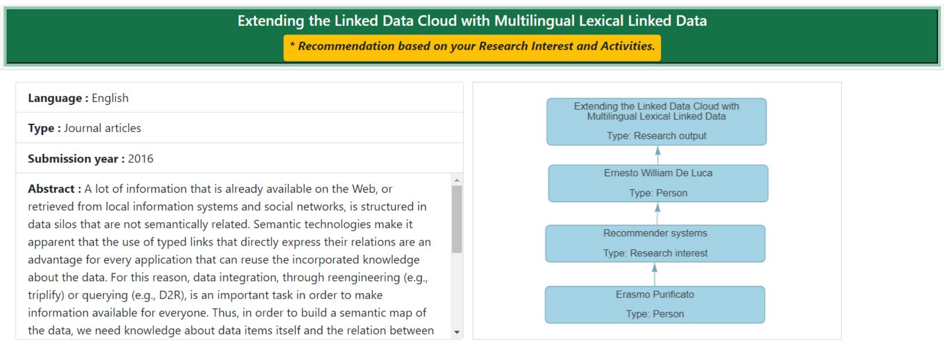


Figure 8: *System A* - Example of KG-path explanation for a particular research output.

- **System B** is composed by a single segment that exhibits the top-10 list of recommended research outputs, as displayed in Fig. 6b. The main explanation here is provided by a *percentage score*, indicating the average similarity score of all the contributions to the recommendation (Fig. 10). Upon clicking on one item, the UI displays the *bar chart* reporting the similarity of the person with the research output on a subgraph level (Fig. 11).

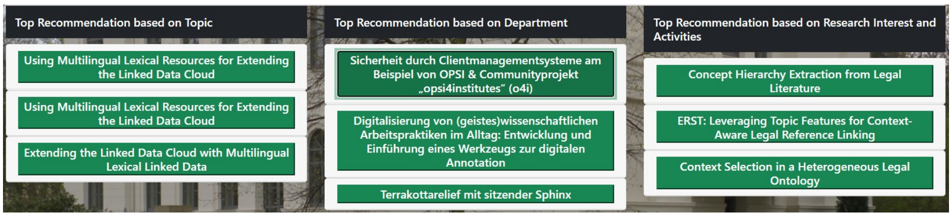


Figure 9: *System A* - Second UI segment displaying the top-3 recommended research outputs for each individual subgraph.

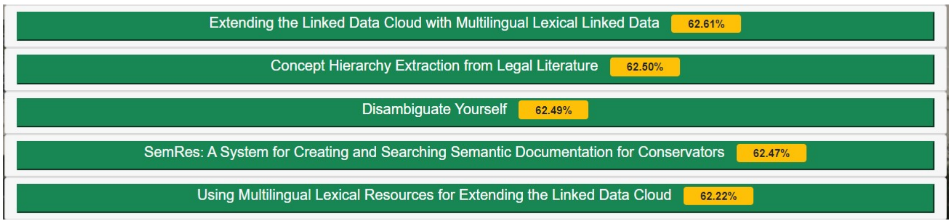


Figure 10: *System B* - Illustrative portion of the UI segment displaying 5 out of the top-10 recommended research outputs.

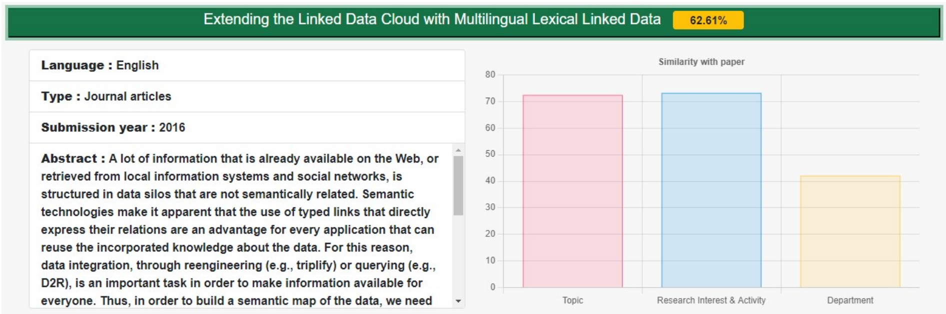


Figure 11: *System B* - Example of graph-chart explanation for a particular research output.

5. Evaluation and results

In order to evaluate the developed graph-based recommender system and the related explainable interfaces, we carried out a user study among the researchers at GEI. They cover several research areas, such as computer science, educational media, linguistics, social sciences and humanities. During the user study, the participants have been asked to solve some guided tasks to gain familiarity with the system and explore the UI to understand all the components. Then, they fill in an online survey to evaluate the effectiveness of both explainable interfaces. The assessment of the explainable interfaces has been performed by using *5-point Likert questions*, where for each of the corresponding statement reported in Table 5, every user gives a response in the range between *Strongly disagree* and *Strongly agree*.

In an ideal scenario, the pool of the participants to the study should have been split into two homogeneous groups, each with the task of assessing only one of the two interfaces (*between-*

Table 5
Online survey questions

Label	System	Question
Q1	A	The explanation sentences were relevant in interpreting the decision-making of the system.
Q1	B	The similarity score was relevant to me in interpreting the decision-making process of the system.
Q2	A	The user-to-paper graph was relevant in understanding the recommendation process.
Q2	B	I found the similarity bar chart relevant in understanding the recommendations.
Q3	A & B	The explanations given by the system help me understand why the items were recommended to me.
Q4	A & B	I understood why the items were recommended to me.
Q5	A & B	I feel confident in using this system as I have attained better understanding of the decision-making process.
Q6	A & B	The explanations have increased my trust in the system.

subjects evaluation), thus making it possible to assess whether the difference in the presentation of the results and their explanation could lead to a different perception of the quality of the recommendations. Due to the small number of available users, each participant first completed the task related to the performance of the recommender and then interacted with both XAI systems (*within-subjects evaluation*). The order of appearance of the two interfaces has been randomly selected for each user to eliminate any bias due to the sequence of interactions that could influence the final evaluation.

The pool of participants is composed of 23 researchers, and the evaluation results are described below. Concerning the assessment of the explainable interfaces, due to the reasons mentioned above related to the number of users available for the study, we ended up doing a qualitative evaluation are displayed in Fig. 12 and Fig. 13. For the visualisation of the Likert scales, we used the *diverging stacked bar charts* as the graphical display technique, based on Robbins and Heiberger's studies on the presentation of results using rating scales [25, 26].

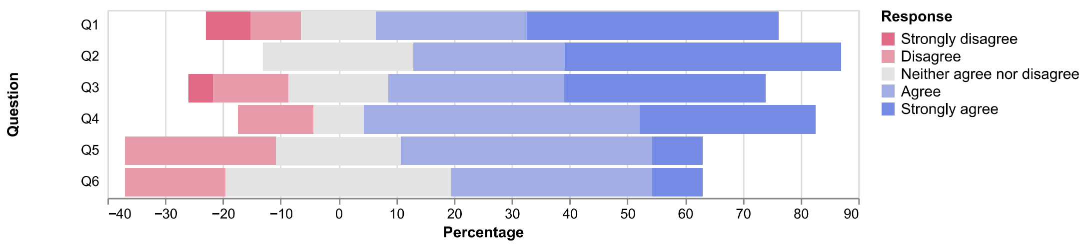


Figure 12: *System A* - Evaluation of the explainable interface

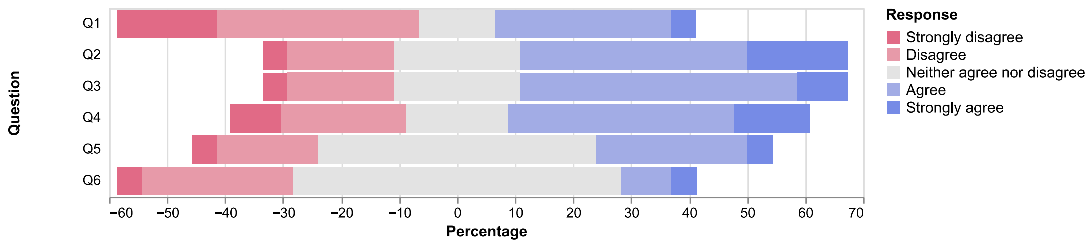


Figure 13: *System B* - Evaluation of the explainable interface

By analysing the result charts, it is clear that the participants expressed their overall preference for *System A* explainable interface. Although the recommendations are pretty understandable for both UIs (Q4), mainly thanks to the explanation (Q3), the greater effectiveness of *System A* compared to *System B* is reported mainly concerning the different types of explanations, namely the one-line-explanation over the similarity score (Q1) and the graph-path over the bar chart (Q2). The questions about *trust* (Q5) and *confidence* (Q6) gained in the system and in using it confirmed that perception.

Finally, to enhance the qualitative evaluation, we conducted some targeted interviews with few users who explicitly agreed to elaborate more about the perceived effectiveness of the two explainable interfaces. Direct questions showed that the differences in judging the two systems are in most cases due to the lack of technical background by humanists, historians and similar fields: "*At first glance, I failed in understanding the meaning of the overall percentage and the charts (in System B), while I found easy to get the motivation behind System A suggestions with explanation in natural language and graph connection*". Some users suggested adding "*the possibility to (explicitly) choose the facets (properties) through which visualise the explanations (e.g. showing a top-10 list made by a combination of only research interests and topics without departments)*". It is also worth reporting a fascinating opinion by a humanist regarding the increase of trust in the system: "*Today I learned: I might not want to know how this system works. I am used to Amazon's recommendations - based on your searches/your purchases/other peoples purchases related to this product. I find this system to be unexpectedly personal. The transparency makes me more aware of what the system 'knows' and 'does'. I would have said I wanted this anytime, but now I see it takes some effort for me to understand and feels a little scary, and also makes me more critical of the quality of the system*". From this insightful evidence, we can derive that a reasonable explanation can "open the eyes" on how an intelligent system works, even to non-technical or sceptical users, leading to a more critical analysis about its effectiveness.

6. Conclusion and future work

In this paper, we presented the implementation of a knowledge graph-based recommender system for research papers developed in the academic context of the Georg Eckert Institute for International Textbook Research by leveraging data from the researchers of the institute and external collaborators. The approach used in the implementation of the presented recommender system is inspired by *entity2rec*, a technique to measure user-item relatedness for top-N item recommendations, which leads to the definition of several subgraphs from the original knowledge graph, considering one relationship between two entities at a time. Besides the recommender, we presented two different explainable interfaces for the visualisation of the results, designed to provide different types of explanations to the users interacting with the implemented recommender, intending to assess which of the different ways of displaying results and related explanations is the most comprehensible and likely to increase confidence and trust in the use of the system by researchers covering different fields of study, such as computer science, educational media, linguistics, social sciences and humanities. By offering explanations in natural language and displaying the graph connections, the results of the evaluation carried out through a user study among the researchers of the GEI showed that *System A* is perceived to be much more effective than *System B* which displays the similarity score between users and papers, along with the bar chart for single property contribution. The analysis of the results themselves and some targeted interviews conducted to enhance the qualitative evaluation suggested further future research activities towards more personalised and adaptive explainability interfaces to let every user gain trust in the recommender system, although they have different demands and technical knowledge. Regarding the development of the recommender system, future activities are planned to exploit more types of entities and relationships in the implementation.

References

[1] M. Pavan, T. Lee, E. W. De Luca, Semantic enrichment for adaptive expert search, in: Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business, Graz, Austria, 2015, pp. 1–4.

[2] V. Mangaravite, R. L. Santos, I. S. Ribeiro, M. A. Gonçalves, A. H. Laender, The lexr collection for expertise retrieval in academia, in: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, Pisa, Italy, 2016, pp. 721–724.

[3] H. A. M. Hassan, Personalized research paper recommendation using deep learning, in: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 2017, pp. 327–330.

[4] X. Bai, M. Wang, I. Lee, Z. Yang, F. Xia, others, Scientific paper recommendation: A survey, IEEE Access 7 (2019) 9324–9339.

[5] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, K.-R. Müller, Explainable AI: interpreting, explaining and visualizing deep learning, volume 11700, Springer Nature, 2019.

[6] N. Tintarev, J. Masthoff, Evaluating the effectiveness of explanations for recommender systems, User Modeling and User-Adapted Interaction 22 (2012) 399–439.

[7] F. Gedikli, D. Jannach, M. Ge, How should i explain? a comparison of different explanation types for recommender systems, International Journal of Human-Computer Studies 72 (2014) 367–382.

[8] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artif. Intell. 267 (2019) 1–38.

[9] M. Millecamp, N. N. Htun, C. Conati, K. Verbert, To explain or not to explain: the effects of

personal characteristics when explaining music recommendations, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, Association for Computing Machinery, Los Angeles, CA, USA, 2019, pp. 397–407.

[10] E. Palumbo, G. Rizzo, R. Troncy, Entity2rec: Learning user-item relatedness from knowledge graphs for top-n item recommendation, in: Proceedings of the eleventh ACM conference on recommender systems, Como, Italy, 2017, pp. 32–36.

[11] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (CSUR) 54 (2021) 1–37.

[12] F. Zhang, N. J. Yuan, D. Lian, X. Xie, W.-Y. Ma, Collaborative knowledge base embedding for recommender systems, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, CA, USA, 2016, pp. 353–362.

[13] Y. Zhang, Q. Ai, X. Chen, P. Wang, Learning over knowledge-base embeddings for recommendation, arXiv preprint arXiv:1803.06540 (2018).

[14] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, M. Guo, Ripplenet: Propagating user preferences on the knowledge graph for recommender systems, in: Proc. of the 27th ACM Int. Conference on Information and Knowledge Manag., Turin, Italy, 2018, pp. 417–426.

[15] Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, Q. He, A survey on knowledge graph-based recommender systems, IEEE Trans. on Knowledge and Data Engin. (2020).

[16] H. Zhao, Q. Yao, J. Li, Y. Song, D. L. Lee, Meta-graph based recommendation fusion over heterogeneous information networks, in: Proc. of the 23rd ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 2017, pp. 635–644.

[17] M. Palmonari, P. Minervini, Knowledge graph embeddings and explainable ai, Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges, IOS Press, Amsterdam (2020) 49–72.

[18] A. Grover, J. Leskovec, node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, CA, USA, 2016, pp. 855–864.

[19] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).

[20] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. T. Riedl, Evaluating collaborative filtering recommender systems, ACM Transactions on Information Systems (TOIS) 22 (2004) 5–53.

[21] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge graph embedding: A survey of approaches and applications, IEEE Trans. on Knowledge and Data Engineering 29 (2017) 2724–2743.

[22] M. Grootendorst, Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics., 2021. URL: https://doi.org/10.5281/zenodo.4381785. doi:10.5281/zenodo.4381785.

[23] L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, arXiv preprint arXiv:1802.03426 (2018).

[24] L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering, Journal of Open Source Software 2 (2017).

[25] N. B. Robbins, R. M. Heiberger, et al., Plotting likert and other rating scales, in: Proceedings of the 2011 Joint Statistical Meeting, Miami, FL, USA, 2011, pp. 1058–1066.

[26] R. M. Heiberger, N. B. Robbins, et al., Design of diverging stacked bar charts for likert scales and other applications, Journal of Statistical Software 57 (2014) 1–32.